

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ (государственный университет)
ФАКУЛЬТЕТ УПРАВЛЕНИЯ И ПРИКЛАДНОЙ МАТЕМАТИКИ
ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР ИМ. А. А. ДОРОДНИЦЫНА РАН
КАФЕДРА «ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ»

Мафусалов Александр Александрович

**Оценивание вероятности успеха в одной серии
испытаний Бернулли по другой серии испытаний при
наличии зависимости между вероятностями успеха**

511656 - Математические и информационные технологии

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Научный руководитель:
с.н.с. ВЦ РАН, д.ф.-м.н.
Воронцов Константин Вячеславович

Москва

2013

Содержание

1	Описание задачи	5
1.1	Область практического применения.	5
1.2	Постановка задачи.	6
2	Подходы к решению задачи.	6
2.1	Верхняя граница функционала \mathcal{L}	6
2.2	Простейшее решение задачи.	8
2.3	Бета распределение как априорное распределение вероятности успеха.	9
2.4	Известные подходы к решению задачи.	10
3	Функциональный подход.	11
3.1	Пропорциональная функция зависимости.	12
3.2	Линейная функция зависимости.	13
3.3	Сигмоидальная функция зависимости.	13
3.4	Функция оценки вероятности успеха во второй серии испытаний.	16
3.5	Линейная функция веса.	17
3.6	Логарифмическая функция веса.	18
3.7	Рациональная функция веса.	18
4	Стохастический подход.	19
4.1	Маргинальные функции распределения.	20
4.2	Копула.	20
4.3	Постановка задачи в рамках стохастического подхода.	21
4.4	Границы Фреше и максимальная копула.	22
4.5	Farli-Gumbel-Morgenstern семейство копул.	25
4.6	Гауссовские копулы.	25
4.7	t копула Стьюдента	26
4.8	Архимедовы копулы.	27
4.9	Семейство Clayton копул.	28
4.10	Семейство Frank копул.	28
4.11	Семейство Gumbel копул.	28
4.12	Вычисление апостериорного распределения.	30
5	Вычислительный эксперимент.	31

5.1	ROC-кривые.	31
5.2	Функциональный подход.	31
5.3	Стохастический подход.	32
6	Заключение	56

Аннотация

В работе рассматривается задача оценивания вероятности успеха в последовательности испытаний Бернулли. Дано множество пар последовательностей испытаний. Требуется построить оценку вероятности успеха в первой последовательности как функцию от количеств испытаний и успехов в двух последовательностях. Предполагается наличие зависимости между вероятностями успеха. В первом рассматриваемом подходе к решению задачи, функциональном, предполагается наличие зависимости между вероятностями успеха. Вводится функция зависимости. Рассматривается параметрическое семейство оценок. Во втором подходе, стохастическом, предполагается наличие априорного совместного распределения для пары вероятностей успеха. Оценка строится как апостериорное матожидание вероятности успеха в первой последовательности испытаний. Совместное распределение задаётся с помощью копул. Оцениваются параметры копул. Производится сравнение двух подходов.

1 Описание задачи

1.1 Область практического применения.

Рассматриваемая задача возникла как подзадача в проблеме предсказания вероятности клика на рекламный баннер пользователем поисковой системы.

Большая часть доходов поисковой системы приходит от коммерческих объявлений. Существует множество рекламодателей, создающих рекламные баннеры, привязанные к некоторым фразам или ключевым словам. Когда пользователь вводит поисковый запрос, система решает, выдать ли в списке результатов рекламу, какую рекламу выдать и где её разместить. Если пользователь заинтересуется рекламным объявлением и кликнет на него, поисковая система получит деньги от рекламодателя за этот клик. Для того, чтобы увеличить свои доходы, поисковая система должна соорудить оптимальный механизм для отбора рекламных объявлений. Учитывается множество факторов, которые можно условно разделить на несколько групп: факторы пользователя (вся информация о пользователе, как то время суток, возраст, пол, история запросов, поисковая система и проч.), факторы запроса (количество слов, наличие определенных ключевых слов), факторы баннера (ключевые слова, история показов). Важную роль в оценке вероятности клика играет статистика показов рекламного баннера — сколько раз этот баннер показывался по различным запросам разным пользователям и сколько раз на баннер кликнули. Больше того, оценка вероятности, полученная как функция показов и кликов, может служить неплохим начальным приближением к общей оценке.

Ранее мы говорили, что поисковая система должна выбрать, где показать объявление. В большинстве систем стоит выбор того, на какой позиции в списке результатов показать объявление, и вероятность клика может существенно зависеть от того, на какой позиции баннер показывается. В поисковой системе Яндекс есть два блока, в которых показывается реклама, так называемые спецразмещение (первые несколько позиций в списке результатов) и гарантия (справа от списка результатов запроса). Вероятности клика у одного и того же баннера существенно отличаются в зависимости от блока показа. При этом, ведётся статистика показов и кликов отдельно по каждому блоку. Возникает ситуация, в которой у баннера есть много показов в гарантии и мало показов в спецразмещении. Но это один и тот же баннер, показанный в разных средах. В данной работе мы хотим построить оценку вероятности

клика в спецразмещении, зависящую от показов и кликов в гарантии.

1.2 Постановка задачи.

Имеется множество объектов $\{x_1, \dots, x_M\}$. Каждому объекту x_i соответствуют две последовательности испытаний Бернулли. Первая имеет длину $n_i + 1$, вероятность успеха p_i и число успехов $k_i + a_i$ (a_i — индикатор успеха $n_i + 1$ -го испытания). Вторая имеет длину m_i , вероятность успеха r_i и число успехов l_i . Предполагается, что вероятности успеха в двух последовательностях связаны некоторой зависимостью. Требуется предсказать вероятность успеха для $n_i + 1$ -го испытания в первой последовательности по результатам предыдущих испытаний в первой последовательности и результатам испытаний во второй последовательности, т.е. построить оценку вероятности успеха как функцию вида $f(n_i, k_i, m_i, l_i)$. Введём обозначение $X_i = (n_i, k_i, m_i, l_i)$. Для измерения качества модели будем использовать функционал правдоподобия. Значение функционала правдоподобия на функции f есть

$$\mathcal{L}(f) = \sum_{i=1}^M a_i \ln f(X_i) + (1 - a_i) \ln(1 - f(X_i)).$$

Пусть задано некоторое параметрическое семейство функций $\mathcal{F}(\Theta)$, такое, что любому $\theta \in \Theta \subset \mathbb{R}^n$ соответствует функция $f_\theta(X) \in \mathcal{F}(\Theta)$, $f_\theta(X) = f(X, \theta) : \mathbb{R}^{n+4} \rightarrow \mathbb{R}$. Тогда исходная задача формулируется как

$$\max_{\theta \in \Theta} \mathcal{L}(f_\theta).$$

2 Подходы к решению задачи.

2.1 Верхняя граница функционала \mathcal{L} .

У рассматриваемой задачи оптимизации есть тривиальное при фиксированной выборке, с одной стороны, и с другой стороны, обеспечивающее глобальный максимум функционала на этой выборке, решение.

Рассмотрим некоторый X и множество неразличимых объектов

$$\Omega(X) = \{x_i | x_i = (X_i, a_i), X_i = X\}.$$

Для этих объектов найдём количество успехов в дополнительных испытаниях

$$c(X) = \sum_{i: X_i = X} a_i,$$

и количество неудач в дополнительных испытаниях

$$s(X) = \sum_{i: X_i=X} (1 - a_i).$$

Так как объекты неразличимы, функция оценки из рассматриваемого семейства возвращает на них одно и то же значение вероятности успеха. С другой стороны, оценки для разных групп неразличимости не обязаны быть связаны между собой. Поэтому задача максимизации правдоподобия редуцируется до совокупности задач на каждой группе неразличимости. Для подмножества $\Omega(X)$ задача запишется следующим образом:

$$\max_{f(X)} \sum_{i: X_i=X} a_i \ln f(X) + (1 - a_i) \ln(1 - f(X)) = \max_{f(X)} c(X) \ln f(X) + s(X) \ln(1 - f(X)).$$

Решением этой задачи является значение

$$f(X) = \frac{c(X)}{c(X) + s(X)},$$

а максимальное значение правдоподобия равно

$$c(X) \ln c(X) + s(X) \ln s(X) - (c(X) + s(X)) \ln(c(X) + s(X)).$$

Таким образом, для зафиксированной выборки строится непараметрическая функция оценки, которую мы будем обозначать $f^*(X)$, обеспечивающая глобальный максимум функционала на этой выборке.

К сожалению, практическое использование такой функции затруднительно по ряду причин:

- Функция не аналитична. Её нельзя записать в виде формулы от четырёх переменных (n, k, m, l) , её описание занимает количество памяти, пропорционально количеству различных объектов в выборке (а количество объектов зависит от задачи и может быть весьма велико).
- Функция не определена для объектов, которых не было в обучающей выборке.
- Функция переобучена. Для того, чтобы понять это, достаточно рассмотреть простой пример. Рассмотрим описание X , для которого $c(X) = 0$. Это означает, что $f^*(X) = 0$. Рассмотрим контрольную подвыборку. Если на ней встретится объект с тем же описанием и успехом в дополнительном испытании, качество

на таком объекте будет $1 \cdot \ln 0 = -\infty$. Следовательно, качество на всей контрольной выборке равно $-\infty$. Остаётся только добавить, что если объектов с указанным типом описания достаточно много, а истинные значения вероятности успеха практически никогда не равны точно 0 или 1, то практически на любой контрольной выборке качество будет равно $-\infty$.

Однако, функция f^* будет полезна для нас как эталон, лучше которого в рамках данной задачи и данной выборки мы ничего получить не сможем.

Целью данной работы является сравнение различных подходов к решению описываемой задачи. Кроме того, однако, нас интересуют функции оценки, существенно использующие информацию о второй серии испытаний. Полезно будет определить редуцированный класс функций оценки \tilde{f} таких, что $\tilde{f}(n, k, m, l) = g(n, k)$. Если зависимость между вероятностями успеха имеет место, то оценки из класса $f(n, k, m, l)$ должны превзойти по качеству оценки из класса $g(n, k)$.

Аналогично тому, как мы построили оценку f^* верхней границы функционала для рассматриваемого изначально класса функций, построим оценку g^* для редуцированного класса функций. Для этого достаточно преобразовать выборку $(X_i, a_i) \rightarrow (Z_i, a_i)$, где $Z_i = (n_i, k_i)$. Далее, определить $c(Z) = \sum_{i:Z_i=X} a_i$, $s(Z) = \sum_{i:Z_i=X} (1 - a_i)$. Тогда оптимальная в классе оценка $g^*(Z) = \frac{c(Z)}{c(Z)+s(Z)}$. Если получится построить такую аналитически представимую функцию $f(X)$, что $\mathcal{L}(f) > \mathcal{L}(g^*)$, то это означает, что мы построили оценку, зависящую от данных первой и второй последовательности более точную, чем это теоретически возможно в классе оценок, зависящих только от данных по первой последовательности.

2.2 Простейшее решение задачи.

Если бы имелся только один объект с серией испытаний Бернулли длиной n с k успехами, то наиболее разумной оценкой, максимизирующей правдоподобие на событиях серии испытаний, была бы оценка $\hat{p} = \frac{k}{n}$.

Такая оценка будет тем точнее, чем больше длина серии испытаний, она является асимптотически несмещённой оценкой вероятности успеха p . Для последовательностей с достаточно большим числом испытаний n эта оценка является достаточно точной. Однако при малых n появление новых испытаний сильно влияет на оценку вероятности успеха.

Мы предполагаем, что в выборке много объектов, а значит есть возможность

оценить априорное распределение вероятностей успеха в первой последовательности и получить оценку вероятности успеха как матожидание вероятности успеха по апостериорному распределению.

2.3 Бета распределение как априорное распределение вероятности успеха.

В задаче прогнозирования вероятности клика в качестве априорного распределения вероятностей клика на множестве баннеров часто используют Бета распределение. Для такого предположения есть ряд причин.

- Плотность Бета распределения с параметрами (α, β) задаётся формулой

$$h(x, \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1},$$

где $B(\alpha, \beta)$ — Бета функция.

Распределение отлично от нуля на отрезке $[0, 1]$, т.е. носитель совпадает с отрезком значений, которые может принимать вероятность успеха.

- Бета распределение является достаточно гибким и с его помощью возможно достаточно точно аппроксимировать широкий класс унимодальных функций распределения. Примеры Бета распределений на рисунке 1
- Если случайная величина p подчиняется Бета распределению $p \propto \text{Beta}(\alpha, \beta)$, а случайные величины χ_1, \dots, χ_n есть независимые Бернуллиевские случайные величины с вероятностью успеха p , то апостериорное распределение величины p есть

$$\bar{p} \propto \text{Beta}\left(\alpha + \sum_i \chi_i, \beta + \sum_i (1 - \chi_i)\right).$$

Оценка вероятности успеха, полученная как матожидание апостериорного распределения

$$\hat{p} = \frac{\sum_i \chi_i + \alpha}{n + \alpha + \beta},$$

является несмещённой оценкой случайной величины p .

- В задаче прогнозирования вероятности клика выберем некоторую длину последовательности n (лучше достаточно большое, например $n = 100$). Вычислим количество объектов с различными значениями k количества успехов $N(k)$ и

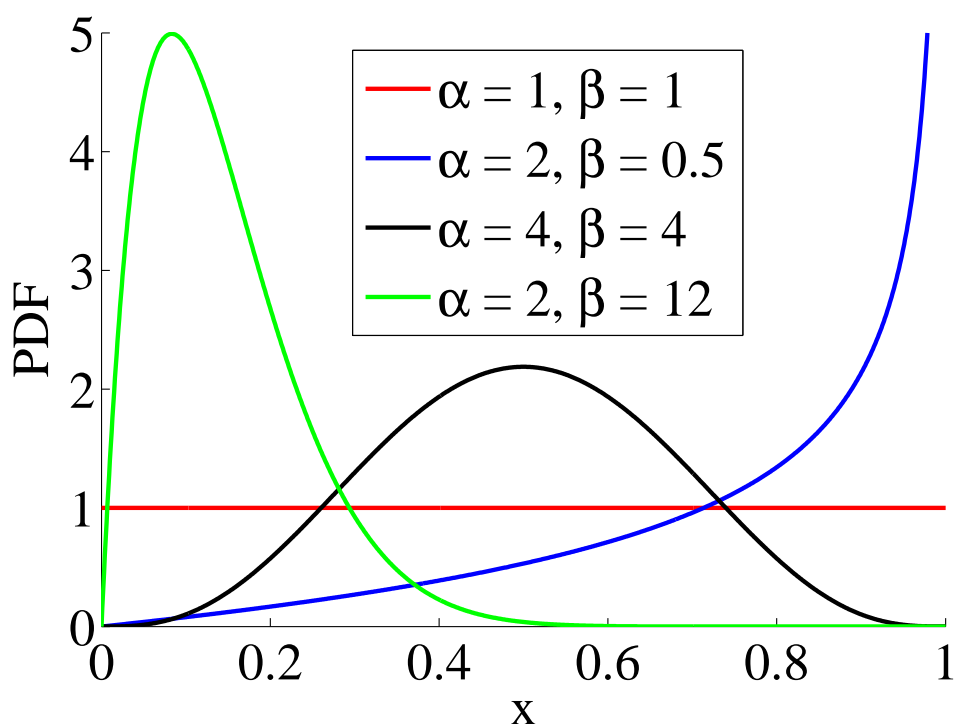


Рис. 1: Графики плотности Бета распределения для различных пар параметров (α, β) .

отложим в координатной плоскости точки $(\frac{k}{n}, N(k))$. Тогда выборка из таких точек может быть достаточно точно аппроксимирована функцией плотности (с точностью до умножения на константу, разумеется) Бета распределения.

- В рассматриваемой задаче возможно подобрать такие параметры (α, β) , что функция оценки

$$f(X) = \frac{k + \alpha}{n + \alpha + \beta},$$

даст прирост качества, по сравнению с простейшей оценкой $f(X) = k/n$, а также неплохое приближение к максимально возможному в данном редуцированном классе оценок качеству, достигаемому на непараметрической оценке $g^*(n, k)$.

2.4 Известные подходы к решению задачи.

На тему учёта эффекта позиции при предсказании вероятности клика на рекламный баннер написано очень много статей.

Как правило, рассматриваются модели, включающие в себя обширный набор факторов: информация о пользователе, запросе, рекламном объявлении, соседних объявлениях [1],[2],[3],[4],[5]. Данная работа ограничивается только одним типом фактора — статистика показов и кликов.

В отдельных работах [3],[6] рассматривается общая статистика показов и кликов, т.е. $(n + m, k + l)$ вместо пары статистик (n, k) и (m, l) .

Для разделённой статистики в работах [2],[1] оценка вероятности строится как линейная комбинация оценок, либо, как в работе [4] — логистическое преобразование от линейной комбинации.

Если, аналогично рассматриваемому в данной работе функциональному подходу, выдвигаются предположения относительно функции зависимости между вероятностями успеха, то чаще всего рассматривается простейшая пропорциональная $(p = \gamma r)$ зависимость [6],[7],[8]. В работе [3] помимо пропорциональной рассматривается аддитивная $(p = r + \gamma)$ зависимость. Оба эти вида зависимости являются частными случаями линейной $(p = \gamma_1 r + \gamma_0)$ зависимости, рассматриваемой в данной работе.

Из стохастических подходов встречается только оценивание методом наивного Байеса [5].

Так же, как и в данной работе, априорные распределения вероятностей успеха описываются при помощи Бета распределений в работах [5],[4],[3].

3 Функциональный подход.

Предположим, что существует общая для всех объектов функция зависимости между вероятностями успеха $p = \phi(r)$. Мы считаем, что вероятности успеха в первой последовательности у объектов, соответствующих описанию X_i изначально распределены согласно Бета распределению с параметрами $(\alpha(X_i), \beta(X_i) - \alpha(X_i))$:

$$p_i^a \propto \text{Beta}(\alpha(X_i), \beta(X_i) - \alpha(X_i)).$$

Тогда ожидаемое значение в апостериорном распределении (с учётом k_i успехов среди n_i испытаний) есть

$$\hat{p}_i = \frac{k_i + \alpha(X_i)}{n_i + \beta(X_i)}.$$

Если бы нам была известна вероятность успеха r_i , мы бы записали $p_i = \phi(r_i)$. Но мы можем предоставить лишь некоторую оценку $\hat{r}_i = r(m_i, l_i)$. Поэтому предположим, что матожидание величины p_i^a равно $\phi(\hat{r}_i)$, откуда следует $\alpha(X_i) = \phi(\hat{r}_i)\beta(X_i)$.

Рассмотрим дисперсию априорного распределения:

$$Dp_i^a = \frac{\alpha(X_i)(\beta(X_i) - \alpha(X_i))}{\beta^2(X_i)(\beta(X_i) + 1)} = \frac{\phi(\hat{r}_i)(1 - \phi(\hat{r}_i))}{\beta(X_i) + 1}.$$

Дисперсия априорной вероятности успеха, с одной стороны, зависит от того, насколько точно мы можем предсказать оценку во второй последовательности \hat{r}_i ; что, в свою очередь зависит преимущественно от длины второй последовательности m_i . С другой стороны, на дисперсию существенно влияет количество испытаний в первой последовательности n_i . Поэтому для регулировки дисперсии будем использовать величины n_i и m_i . Из этих соображений будем рассматривать функции вида $\beta(X_i) = \mu(n_i, m_i)$. Будем называть функцию μ функцией веса, т.к. она позволяет взвесить влияние прямой статистики из первой последовательности испытаний (числа k и n) и косвенной статистики из второй последовательности испытаний (числа l и m).

В итоге, получаем следующий вид функции оценки:

$$\hat{p}_i = \frac{k_i + \phi(r(m_i, l_i))\mu(n_i, m_i)}{n_i + \mu(n_i, m_i)}.$$

Пусть теперь выбраны параметрические семейства $r(m_i, l_i, \theta)$, $\phi(\hat{r}, \gamma)$, $\mu(n_i, m_i, \lambda)$. Тогда функция оценки представляется как функция $f(X, \theta, \gamma, \lambda)$, а исходная задача формулируется как

$$\mathcal{L}(f(X, \theta, \gamma, \lambda)) \rightarrow \max_{\theta, \gamma, \lambda}.$$

3.1 Пропорциональная функция зависимости.

Простейшим предположением является то, что функция зависимости имеет вид $\phi(\hat{r}, \gamma) = \gamma\hat{r}$. Т.е. предполагается, что отношение вероятностей успеха в двух сериях постоянно для всех объектов. Обозначим $\mathcal{L}^{(1)}(a, f(X)) = a \ln f(X) + (1 - a) \ln(1 - f(X))$. В этом случае производная:

$$\frac{\partial \mathcal{L}^{(1)}(a, f(X, \theta, \gamma, \lambda))}{\partial \gamma} = \frac{\partial \mathcal{L}^{(1)}}{\partial f} \frac{\partial f}{\partial \phi} \frac{\partial \phi}{\partial \gamma},$$

где

$$\frac{\partial \mathcal{L}^{(1)}}{\partial f} = \frac{\partial}{\partial f} [(a \ln f(X) + (1 - a) \ln(1 - f(X)))] = \frac{a}{f(X)} - \frac{1 - a}{1 - f(X)},$$

$$\frac{\partial f}{\partial \phi} = \frac{\partial}{\partial \phi} \frac{k + \phi\mu}{n + \mu} = \frac{\mu(n, m)}{n + \mu(n, m)},$$

$$\frac{\partial \phi}{\partial \gamma} = \hat{r} = r(m, l).$$

Таким образом, производная функционала правдоподобия по параметру γ равна

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \gamma} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\mu(n_i, m_i, \lambda)}{n_i + \mu(n_i, m_i, \lambda)} r(m_i, l_i, \theta).$$

3.2 Линейная функция зависимости.

Естественным обобщением простейшей пропорциональной зависимости является линейная зависимость: $\phi(\hat{r}, \gamma) = \gamma_1 \hat{r} + \gamma_0$. Воспользовавшись выкладками из предыдущего пункта, получим формулы для производных:

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \gamma_1} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\mu(n_i, m_i, \lambda)}{n_i + \mu(n_i, m_i, \lambda)} r(m_i, l_i, \theta),$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \gamma_0} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\mu(n_i, m_i, \lambda)}{n_i + \mu(n_i, m_i, \lambda)}.$$

3.3 Сигмоидальная функция зависимости.

У двух предыдущих функций, несмотря на простоту вычисления и интерпретируемость, есть существенный недостаток: они, вообще говоря, не являются отображением отрезка $[0, 1]$ на себя. Поэтому введем семейство функций зависимости, удовлетворяющее следующим свойствам:

- $0 < \phi(r) < 1$ для $0 < r < 1$.
- $\phi(0) = 0$, $\phi(1) = 1$.
- $\phi'(r) > 0$.

Эти требования отражают естественные соображения из предметной области:

- Вероятность успеха в первой последовательности должна быть интерпретируема как вероятность успеха, т.е. быть числом между 0 и 1.
- Если на баннер вообще никогда не кликнул в одном блоке, то на него никогда не кликнул и в другом блоке. И наоборот: если на баннер всегда кликают в одном блоке, то на него всегда будут кликать и в другом блоке.
- Если один баннер лучше другого в одном блоке, то он лучше и в другом блоке.

Рассмотрим параметрическое семейство функций

$$\phi(r, \gamma) = \exp \left\{ -\frac{(1-r)(\gamma_1 r + \gamma_0(1-r))}{r} \right\},$$

где $\gamma_0, \gamma_1 > 0$.

Покажем, что функции $\phi(r, \gamma)$ удовлетворяют указанным выше требованиям:

- Первое требование, как нетрудно заметить, следует из второго и третьего, поэтому докажем их.

•

$$(1-r)(\gamma_1 r + \gamma_0(1-r)) \rightarrow \gamma_0 > 0,$$

при $r \rightarrow 0$, следовательно,

$$-\frac{(1-r)(\gamma_1 r + \gamma_0(1-r))}{r} \rightarrow -\infty,$$

при $r \rightarrow 0$, следовательно, $\phi(0, \gamma) = e^{-\infty} = 0$.

$$\phi(1, \gamma) = \exp \left\{ -\frac{0 \cdot \gamma_1}{1} \right\} = 1.$$

- Рассмотрим производную

$$\begin{aligned} & \frac{\partial}{\partial r} \frac{(1-r)(\gamma_1 r + \gamma_0(1-r))}{r} = \\ & = \frac{-r(\gamma_1 r + \gamma_0(1-r)) + \gamma_1(1-r)r - \gamma_0(1-r)r - (1-r)(\gamma_1 r + \gamma_0(1-r))}{r^2} = \\ & = \frac{r^2(-\gamma_1 + \gamma_0 - \gamma_1 + \gamma_0 + \gamma_1 - \gamma_0) + r(-\gamma_0 + \gamma_1 - \gamma_0 - \gamma_1 + \gamma_0 + \gamma_0) - \gamma_0}{r^2} = \\ & = \frac{(r^2 - 1)\gamma_0 + (-r^2)\gamma_1}{r^2}. \end{aligned}$$

Заметим, что $\phi'_r(r, \gamma) > 0 \Leftrightarrow (r^2 - 1)\gamma_0 + (-r^2)\gamma_1 < 0$. При $r \rightarrow 0$ получим требование $\gamma_0 > 0$. При $r \rightarrow 1$ получим требование $\gamma_1 > 0$. С другой стороны, $\gamma_0, \gamma_1 > 0 \Rightarrow (r^2 - 1)\gamma_0 + (-r^2)\gamma_1 < 0$. Следовательно,

$$\phi'_r(r, \gamma) > 0 \Leftrightarrow \gamma_0, \gamma_1 > 0.$$

Найдём производные функции ϕ по параметрам:

$$\frac{\partial \phi}{\partial \gamma_0} = -\frac{(1-r)^2}{r} \phi(r, \gamma),$$

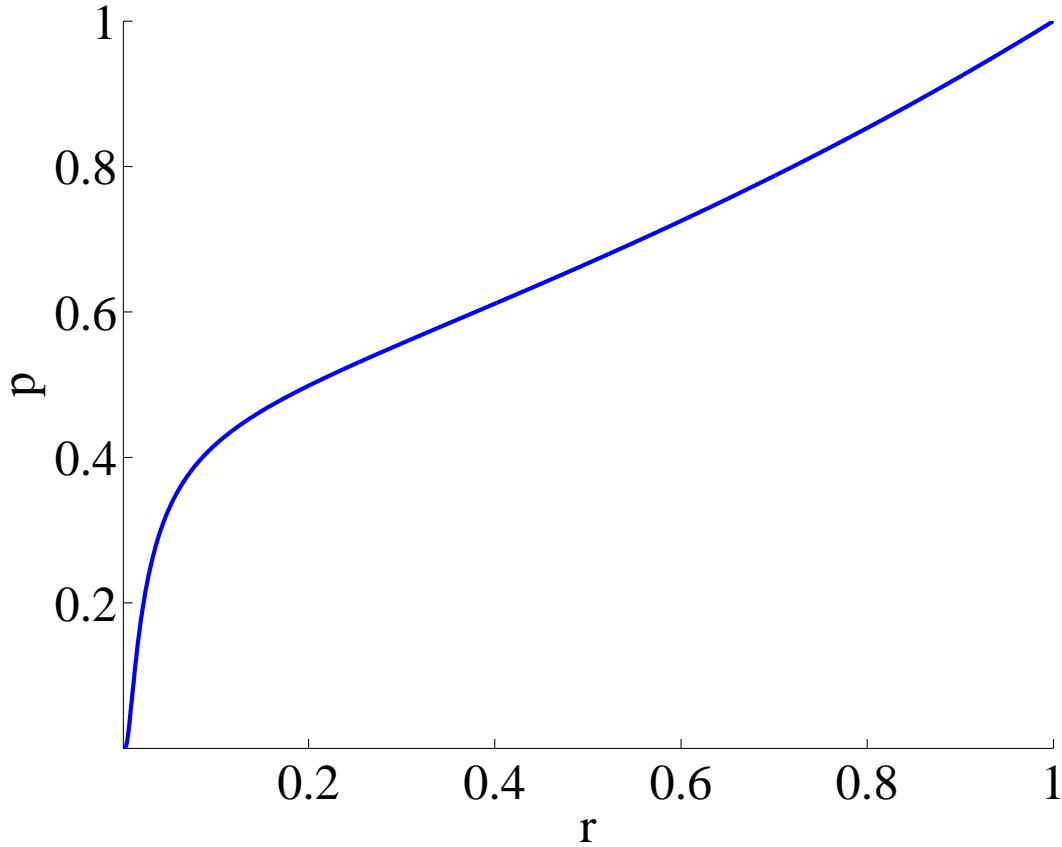


Рис. 2: Сигмоидальная функция зависимости с параметрами $\gamma_0 = 0.79$, $\gamma_1 = 0.02$.

$$\frac{\partial \phi}{\partial \gamma_0} = -(1-r)\phi(r, \gamma).$$

Поэтому производные функционала правдоподобия:

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \gamma_0} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1-a_i}{1-f(X_i, \theta, \gamma, \lambda)} \right) \frac{\mu(n_i, m_i, \lambda)}{n_i + \mu(n_i, m_i, \lambda)}.$$

$$\cdot \left(-\frac{(1-r(m_i, l_i, \theta))^2}{r(m_i, l_i, \theta)} \right) \phi(r(m_i, l_i, \theta), \gamma),$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \gamma_1} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1-a_i}{1-f(X_i, \theta, \gamma, \lambda)} \right) \frac{\mu(n_i, m_i, \lambda)}{n_i + \mu(n_i, m_i, \lambda)}.$$

$$\cdot (r(m_i, l_i, \theta) - 1)\phi(r(m_i, l_i, \theta), \gamma).$$

Пример сигмоидальной зависимости на рис. 2.

3.4 Функция оценки вероятности успеха во второй серии испытаний.

Рассмотрим случайную величину \hat{r}^a — априорная вероятность успеха во второй последовательности испытаний. Если не накладывать никаких ограничений на распределение \hat{r}^a , то при условии, что из m испытаний было l успехов, оценка максимума правдоподобия есть

$$\hat{r}_{ML}(m, l) = \frac{l}{m}.$$

Предположим, что априорное распределение есть Бета-распределение с параметрами $(\alpha, \beta - \alpha)$: $\hat{r}^a \propto \text{Beta}(\alpha, \beta - \alpha)$. Тогда, если из m испытаний было l успехов, то апостериорная вероятность успеха имеет распределение $\hat{r}^p \propto \text{Beta}(\alpha + l, (\beta + m) - (\alpha + l))$. Поэтому ожидаемая вероятность успеха:

$$\hat{r}_{\alpha, \beta}(m, l) = \frac{l + \alpha}{m + \beta}.$$

Примем в качестве семейства функций оценки вероятности успеха во второй последовательности испытаний функции $r(m, l, \theta) = \frac{l + \theta_0}{m + \theta_1}$, где $\theta_0, \theta_1 > 0$. Вычислим производные:

$$\begin{aligned} \frac{\partial r}{\partial \theta_0} &= \frac{1}{m + \theta_1}, \\ \frac{\partial r}{\partial \theta_1} &= -\frac{l + \theta_0}{(m + \theta_1)^2}. \end{aligned}$$

$$\frac{\partial \mathcal{L}^{(1)}(a, f(X, \theta, \gamma, \lambda))}{\partial \theta_j} = \frac{\partial \mathcal{L}^{(1)}}{\partial f} \frac{\partial f}{\partial \phi} \frac{\partial \phi}{\partial r} \frac{\partial r}{\partial \theta_j}.$$

Для каждого из описанных выше семейств функций зависимости найдем вид производной $\phi'_r(r, \gamma)$ и выпишем производные функционала правдоподобия по параметрам θ_j :

- Пропорциональная зависимость.

$$\phi(r, \gamma) = \gamma r, \quad \phi'_r(r, \gamma) = \gamma,$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \theta_0} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\mu(n_i, m_i, \lambda)}{n_i + \mu(n_i, m_i, \lambda)} \gamma \frac{1}{m_i + \theta_1},$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \theta_1} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\mu(n_i, m_i, \lambda)}{n_i + \mu(n_i, m_i, \lambda)} \gamma \left(-\frac{l_i + \theta_0}{(m_i + \theta_1)^2} \right).$$

- Линейная зависимость.

$$\phi(r, \gamma) = \gamma_1 r + \gamma_0, \quad \phi'_r(r, \gamma) = \gamma_1,$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \theta_0} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\mu(n_i, m_i, \lambda)}{n_i + \mu(n_i, m_i, \lambda)} \gamma_1 \frac{1}{m_i + \theta_1},$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \theta_0} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\mu(n_i, m_i, \lambda)}{n_i + \mu(n_i, m_i, \lambda)} \gamma_1 \left(-\frac{l_i + \theta_0}{(m_i + \theta_1)^2} \right).$$

- Сигмоидальная зависимость.

$$\phi(r, \gamma) = \exp \left\{ -\frac{(1-r)(\gamma_1 r + \gamma_0(1-r))}{r} \right\},$$

$$\phi'_r(r, \gamma) = -\frac{r^2(\gamma_0 - \gamma_1) - \gamma_0}{r^2} \phi(r, \gamma) = \left((\gamma_1 - \gamma_0) + \frac{\gamma_0}{r^2} \right) \phi(r, \gamma),$$

$$\begin{aligned} \frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \theta_0} &= \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\mu(n_i, m_i, \lambda)}{n_i + \mu(n_i, m_i, \lambda)} \\ &\cdot \left((\gamma_1 - \gamma_0) + \frac{\gamma_0}{r(m_i, l_i, \theta)^2} \right) \phi(r(m_i, l_i, \theta), \gamma) \frac{1}{m_i + \theta_1}, \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \theta_0} &= \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\mu(n_i, m_i, \lambda)}{n_i + \mu(n_i, m_i, \lambda)} \\ &\cdot \left((\gamma_1 - \gamma_0) + \frac{\gamma_0}{r(m_i, l_i, \theta)^2} \right) \phi(r(m_i, l_i, \theta), \gamma) \left(-\frac{l_i + \theta_0}{(m_i + \theta_1)^2} \right). \end{aligned}$$

3.5 Линейная функция веса.

В качестве простейшего семейства функций веса можно рассмотреть семейство линейных функций

$$\mu(n, m, \lambda) = \lambda_2 n + \lambda_1 m + \lambda_0.$$

В таком случае

$$\frac{\partial \mu}{\partial \lambda_2} = n, \quad \frac{\partial \mu}{\partial \lambda_1} = m, \quad \frac{\partial \mu}{\partial \lambda_0} = 1,$$

$$\frac{\partial \mathcal{L}^{(1)}(a, f(X, \theta, \gamma, \lambda))}{\partial \lambda_j} = \frac{\partial \mathcal{L}^{(1)}}{\partial f} \frac{\partial f}{\partial \mu} \frac{\partial \mu}{\partial \lambda_j},$$

$$\frac{\partial f}{\partial \mu} = \frac{\phi(m + \mu) - (k + \phi\mu)}{(m + \mu)^2} = \frac{\phi m - k}{(m + \mu)^2},$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \lambda_0} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\phi(r(m_i, l_i, \theta), \gamma) m_i - k_i}{(m_i + \mu(n_i, m_i))^2},$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \lambda_1} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\phi(r(m_i, l_i, \theta), \gamma) m_i - k_i}{(m_i + \mu(n_i, m_i))^2} m_i,$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \lambda_2} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\phi(r(m_i, l_i, \theta), \gamma) m_i - k_i}{(m_i + \mu(n_i, m_i))^2} n_i,$$

3.6 Логарифмическая функция веса.

Заметим, что чем больше испытаний во второй последовательности, тем точнее оценка \hat{r} и, следовательно, точнее оценка $\phi(\hat{r})$. Поэтому потребуем роста функции $\mu(n, m)$ при росте m .

С другой стороны, чем больше n , тем точнее оценка $\hat{p} = \frac{k}{n}$, следовательно, оценка

$$f(X) = \frac{\frac{k}{n} + \frac{\phi \mu}{n}}{1 + \frac{\mu}{n}} \rightarrow \frac{k}{n}, \quad n \rightarrow \infty,$$

$$\Rightarrow \frac{\mu}{n} \rightarrow 0, \quad n \rightarrow \infty.$$

В качестве семейства, удовлетворяющего этим двум условиям, можно взять

$$\mu(n, m, \lambda) = \lambda_2 \ln(n + 1) + \lambda_1 \ln(m + 1) + \lambda_0,$$

$$\frac{\partial \mu}{\partial \lambda_2} = \ln(n + 1), \quad \frac{\partial \mu}{\partial \lambda_1} = \ln(m + 1), \quad \frac{\partial \mu}{\partial \lambda_0} = 1,$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \lambda_0} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\phi(r(m_i, l_i, \theta), \gamma) m_i - k_i}{(m_i + \mu(n_i, m_i))^2},$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \lambda_1} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\phi(r(m_i, l_i, \theta), \gamma) m_i - k_i}{(m_i + \mu(n_i, m_i))^2} \ln(m_i + 1),$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \lambda_2} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\phi(r(m_i, l_i, \theta), \gamma) m_i - k_i}{(m_i + \mu(n_i, m_i))^2} \ln(n_i + 1).$$

3.7 Рациональная функция веса.

Если функция $\mu(n, m)$ неограниченно возрастает при росте m , может появиться следующая проблема: пусть m велико, тогда $\mu(n, m) \gg n$, а объект отклоняется от найденной зависимости $p = \phi(r)$. Тогда, даже при достаточно большом n , при котором достаточно точна оценка $\hat{p} = \frac{k}{n}$, оценка $\hat{p} = \frac{k + \phi \mu}{n + \mu}$ будет значительно отклоняться от $\hat{p} = \frac{k}{n}$. Этой ситуации можно избежать, если потребовать ограниченности функции веса. Поэтому дополнительно к требованиям, сделанным в предыдущем пункте, введём следующее: $\mu(n, m, \lambda) \leq \lambda_0$. Рассмотрим следующее семейство функций

$$\mu(n, m, \lambda) = \frac{\lambda_0 m}{(\lambda_2 n + 1)(m + \lambda_1)}, \quad \lambda_0, \lambda_2 > 0,$$

которое, как несложно убедиться, удовлетворяет следующим требованиям:

- $\mu(n, m) \nearrow$ при $m \nearrow$,
- $\mu(n, m) \searrow$ при $n \nearrow$,
- для некоторого C выполнено $\mu(n, m) \leq C$.

Вычислим производные:

$$\frac{\partial \mu}{\partial \lambda_0} = \frac{m}{(\lambda_2 n + 1)(m + \lambda_1)}, \quad \frac{\partial \mu}{\partial \lambda_1} = -\frac{\mu}{m + \lambda_1}, \quad \frac{\partial \mu}{\partial \lambda_2} = -\frac{\mu n}{\lambda_2 n + 1},$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \lambda_0} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\phi(r(m_i, l_i, \theta), \gamma) m_i - k_i}{(m_i + \mu(n_i, m_i))^2} \cdot \frac{m_i}{(\lambda_2 n_i + 1)(m_i + \lambda_1)},$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \lambda_1} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\phi(r(m_i, l_i, \theta), \gamma) m_i - k_i}{(m_i + \mu(n_i, m_i))^2} \cdot \frac{(-\mu(n_i, m_i, \lambda))}{m_i + \lambda_0},$$

$$\frac{\partial \mathcal{L}(f(X, \theta, \gamma, \lambda))}{\partial \lambda_2} = \sum_{i=1}^M \left(\frac{a_i}{f(X_i, \theta, \gamma, \lambda)} - \frac{1 - a_i}{1 - f(X_i, \theta, \gamma, \lambda)} \right) \frac{\phi(r(m_i, l_i, \theta), \gamma) m_i - k_i}{(m_i + \mu(n_i, m_i))^2} \cdot \frac{(-\mu(n_i, m_i, \lambda) n_i)}{\lambda_2 n_i + 1}.$$

4 Стохастический подход.

В этой главе мы наложим несколько иное предположение на природу зависимости между вероятностями успеха p и r . Пусть у множества объектов есть начальное двумерное априорное распределение $H(p, r)$ с плотностью $h(p, r)$.

Тогда матожидание апостериорной вероятности успеха

$$E(p|X) = \frac{1}{Z} \int \int p v(p, r) h(p, r) dp dr,$$

где

$$v(p, r) \cdot C = P(X|p, r) = p^{k_i} (1 - p)^{n_i - k_i} C_{n_i}^{k_i} r^{l_i} (1 - r)^{m_i - l_i} C_{m_i}^{l_i}, \quad C = C_{n_i}^{k_i} C_{m_i}^{l_i},$$

Z — нормировочная константа:

$$Z = \int \int v(p, r) h(p, r) dp dr.$$

В качестве оценки вероятности успеха берётся это матожидание:

$$f(X) = E(p|X).$$

Таким образом, для того, чтобы вычислить значение функции $f(X)$ достаточно определить двумерную функцию распределения $H(p, r)$.

4.1 Маргинальные функции распределения.

Как уже говорилось ранее, предположение о том, что априорные вероятности успеха подчинены Бета-распределению даёт неплохую начальную оценку вероятности успеха. Поэтому потребуем, чтобы маргинальные распределения $H_1(p)$ и $H_2(r)$ искомого двумерного распределения $H(p, r)$ были Бета-распределениями, тогда

$$h_1(p, \xi) = \frac{1}{B(\xi_0, \xi_1)} p^{\xi_0-1} (1-p)^{\xi_1-1},$$

$$h_2(r, \zeta) = \frac{1}{B(\zeta_0, \zeta_1)} p^{\zeta_0-1} (1-p)^{\zeta_1-1},$$

где $h_1(p, \xi)$ и $h_2(r, \zeta)$ — плотности маргинальных распределений $H_1(p)$ и $H_2(r)$, $B(\alpha, \beta)$ — Бета-функция.

Для того, чтобы смоделировать удовлетворяющее заданным требованиям распределение, нам потребуется ввести понятие копулы.

4.2 Копула.

Определение 1 Копулой назовём многомерную функцию распределения, определённую на n -мерном единичном кубе $[0, 1]^n$, такую что каждое её маргинальное распределение равномерно на интервале $[0, 1]$.

Теорема 1 (Склара) Пусть задано многомерное распределение $H(x_1, \dots, x_n)$ с маргинальными распределениями $H_1(x_1), \dots, H_n(x_n)$. Тогда

- Существует копула C такая, что

$$H(x_1, \dots, x_n) = (H_1(x_1), \dots, H_n(x_n)).$$

- Если распределения $H_1(x_1), \dots, H_n(x_n)$ непрерывны, то существует единственная копула C такая, что $H(x_1, \dots, x_n) = (H_1(x_1), \dots, H_n(x_n))$.

Копула, таким образом, предоставляет информацию о том, как связаны случайные величины между собой, игнорируя их индивидуальные распределения. Многие известные меры зависимости между случайными величинами являются оценками некоторой функции от копулы распределения.

Рассмотрим коэффициент ранговой корреляции Спирмена:

$$\rho_n = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 \sum_{i=1}^n (S_i - \bar{S})^2}} \in [-1; 1],$$

где (R_i, S_i) — пара рангов, соответствующих величинам (x_i, y_i) ,

$$\bar{R} = \frac{1}{n} \sum_{i=1}^n R_i = \frac{n+1}{2} = \frac{1}{n} \sum_{i=1}^n S_i = \bar{S}.$$

Оказывается [9], что ρ_n — асимптотически несмещенная оценка величины

$$\rho = 12 \int_0^1 \int_0^1 u_1 u_2 dC(u_1, u_2) - 3 = 12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3.$$

Рассмотрим коэффициент ранговой корреляции Кендалла:

$$\tau_n = \frac{P_n - Q_n}{C_n^2},$$

где P_n и Q_n — число конкордантных и дискордантных пар соответственно. Две пары (x_i, y_i) и (x_j, y_j) называются конкордантными, если $(x_i - x_j)(y_i - y_j) > 0$ и дискордантными если $(x_i - x_j)(y_i - y_j) < 0$ (вероятность события $(x_i - x_j)(y_i - y_j) = 0$ считаем нулевой ввиду непрерывности рассматриваемых распределений).

Оказывается [9], что τ_n — асимптотически несмещенная оценка величины

$$\tau = 4 \int_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1.$$

4.3 Постановка задачи в рамках стохастического подхода.

Теорема Склара позволит нам моделировать распределение $H(p, r)$ следующим образом: пусть $C(u_1, u_2, \eta)$ — некоторое параметрическое семейство копул. Будем искать функцию H в виде

$$H(p, r, \xi, \zeta, \eta) = C(H_1(p, \xi), H_2(r, \zeta), \eta),$$

где $H_1(p, \xi)$ и $H_2(r, \zeta)$ — функции Бета распределения.

Обозначим за $c(u_1, u_2, \eta)$ двумерную плотность распределения копулы. Тогда функция оценки имеет вид:

$$f(X, \xi, \zeta, \eta) = \frac{1}{Z} \int_0^1 \int_0^1 p v(p, r) h_1(p, \xi) h_2(r, \zeta) c(H_1(p, \xi), H_2(r, \zeta), \eta) dp dr,$$

где

$$Z = \int_0^1 \int_0^1 v(p, r) h_1(p, \xi) h_2(r, \zeta) c(H_1(p, \xi), H_2(r, \zeta), \eta) dp dr.$$

Исходная задача формулируется как

$$\mathcal{L}(f(X, \xi, \zeta, \eta)) \rightarrow \max_{\xi, \zeta, \eta}.$$

4.4 Границы Фреше и максимальная копула.

Теорема 2 (Фреше) Для любой копулы $C : [0, 1]^d \rightarrow [0, 1]$ и любых $(u_1, \dots, u_d) \in [0, 1]^d$ выполнены следующие неравенства:

$$W(u_1, \dots, u_d) \leq C(u_1, \dots, u_d) \leq M(u_1, \dots, u_d),$$

где

$$W(u_1, \dots, u_d) = \max \left\{ 1 - d + \sum_{i=1}^d u_i, 0 \right\},$$

$$M(u_1, \dots, u_d) = \min\{u_1, \dots, u_d\}.$$

Ясно, что для размерности 2 неравенство приобретает вид:

$$\max\{u_1 + u_2 - 1, 0\} \leq C(u_1, u_2) \leq \min\{u_1, u_2\}.$$

Графики с минимальной и максимальной копулами на рисунках 3,4.

Максимальная и минимальная копулы W и M отражают два крайних случая зависимости между переменными: если существует детерминистическая зависимость $U_1 = \phi(U_2)$ и функция ϕ — монотонно возрастающая, то копула, соответствующая совместному распределению будет максимальной копулой.

Теорема 3 Пусть для случайных величин U_1 и U_2 выполнено: $U_1 = \phi(U_2)$, и ϕ — непрерывная возрастающая функция. Тогда копула совместного распределения величин U_1 и U_2 есть

$$C(u_1, u_2) = \min\{u_1, u_2\}.$$

Доказательство. Пусть F_1 и F_2 — функции распределения величин U_1 и U_2 . Тогда

$$F_1(x) = P(U_1 \leq x) = P(\phi(U_2) \leq x) = P(U_2 \leq \phi^{-1}(x)) = F_2(\phi^{-1}(x)),$$

$$F_1(\phi(x)) = F_2(x), \quad F_2^{-1}(x) = \phi^{-1}(F_1^{-1}(x)),$$

$$C(u_1, u_2) = P(U_1 \leq F_1^{-1}(u_1), U_2 \leq F_2^{-1}(u_2)) = P(U_2 \leq \phi^{-1}(F_1^{-1}(u_1)), U_2 \leq F_2^{-1}(u_2)) =$$

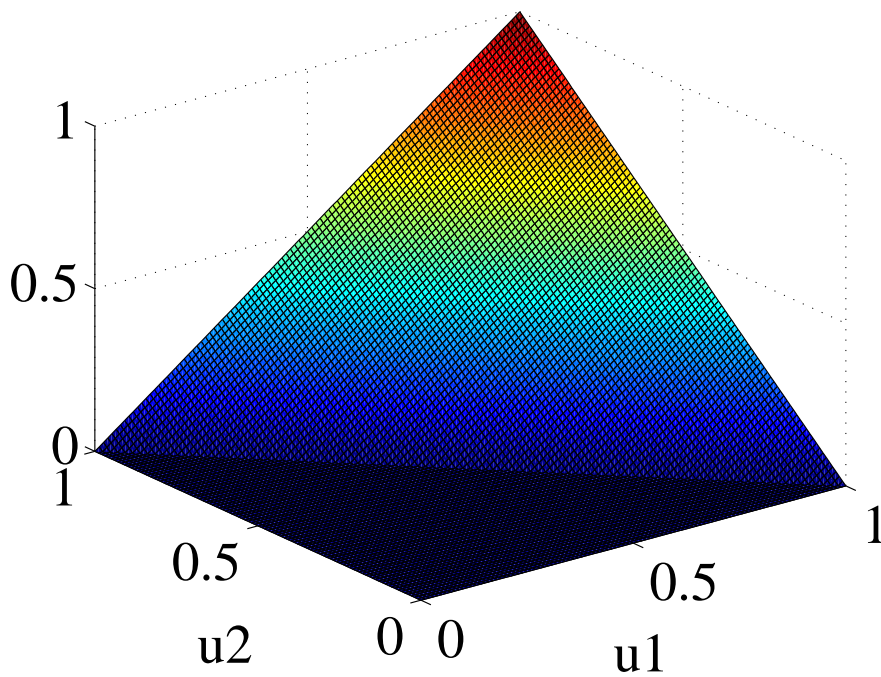


Рис. 3: Минимальная копула $W(u_1, \dots, u_d) = \max \{1 - d + \sum_{i=1}^d u_i, 0\}$.

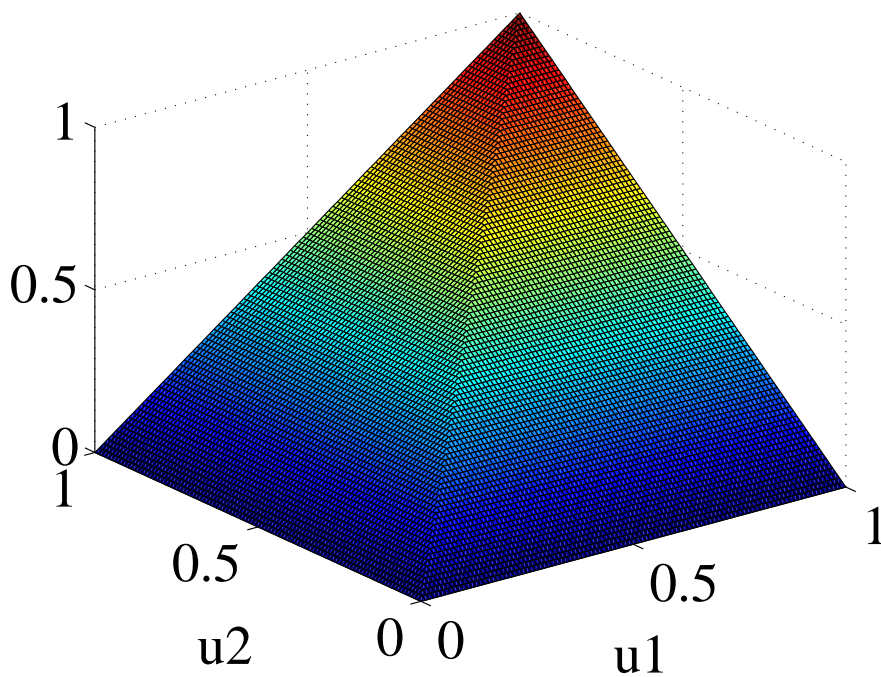


Рис. 4: Максимальная копула $M(u_1, \dots, u_d) = \min\{u_1, \dots, u_d\}$.

$$= P(U_2 \leq F_2^{-1}(u_1), U_2 \leq F_2^{-1}(u_2)) = F_2(\min\{F_2^{-1}(u_1), F_2^{-1}(u_2)\}) = \min\{u_1, u_2\}.$$

Покажем теперь, что аналогичное утверждение выполнено для убывающей функции зависимости и минимальной копулы.

Теорема 4 Пусть для случайных величин U_1 и U_2 выполнено: $U_1 = \phi(U_2)$, и ϕ — непрерывная убывающая функция. Тогда копула совместного распределения величин U_1 и U_2 есть

$$C(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}.$$

Доказательство. Пусть F_1 и F_2 — функции распределения величин U_1 и U_2 . Тогда

$$F_1(x) = P(U_1 \leq x) = P(\phi(U_2) \leq x) = P(U_2 \geq \phi^{-1}(x)) = 1 - F_2(\phi^{-1}(x)),$$

$$F_2(x) = 1 - F_1(\phi(x)), \quad F_2^{-1}(x) = \phi^{-1}(F_1^{-1}(1 - x)), \quad F_2^{-1}(1 - x) = \phi^{-1}(F_1^{-1}(x)),$$

$$\begin{aligned} C(u_1, u_2) &= P(U_1 \leq F_1^{-1}(u_1), U_2 \leq F_2^{-1}(u_2)) = P(U_2 \geq \phi^{-1}(F_1^{-1}(u_1)), U_2 \leq F_2^{-1}(u_2)) = \\ &= P(U_2 \geq F_2^{-1}(1 - u_1), U_2 \leq F_2^{-1}(u_2)) = \\ &= \{0, 1 - u_1 > u_2; F_2(F_2^{-1}(u_2)) - F_2(F_2^{-1}(1 - u_1)), 1 - u_1 < u_2\} = \max\{u_1 + u_2 - 1, 0\}. \end{aligned}$$

Теорема 5 Если копула совместного распределения является максимальной копулой или минимальной копулой, то она остаётся неизменной при переходе от априорного распределения к апостериорному.

Доказательство. Пусть копула максимальна. Это означает, что $P(F_1(U_1) > F_2(U_2)) = 0$ и $P(F_1(U_1) < F_2(U_2)) = 0$, т.е. $U_1 = F_1^{-1}(F_2(U_2)) = \phi(U_2)$ (заметим, что ϕ — возрастающая функция). Апостериорное распределение есть условная вероятность $P(U_1 \leq u_1, U_2 \leq u_2 | A(U_1, U_2))$. Но при добавлении условия области меры 0 таковыми останутся: $P(U_1 > \phi(U_2) | A(U_1, U_2)) = 0$ и $P(U_1 < \phi(U_2) | A(U_1, U_2)) = 0$. Поэтому сохранится утверждение $\tilde{U}_1 = \phi(\tilde{U}_2)$, притом функция ϕ осталась прежней (непрерывной и возрастающей), следовательно совместному распределению величин \tilde{U}_1 и \tilde{U}_2 соответствует максимальная копула. Аналогично проводится доказательство и для минимальной копулы.

Максимальная копула особенно интересна в связи с предположениями функционального подхода. Здесь, в стохастическом подходе, предположение $p = \phi(r)$ автоматически приводит нас к выводу о том, что совместному распределению величин p и r соответствует максимальная копула, и для определения функции оценки $f(X)$

остаётся только задать два маргинальных распределения и вычислить ожидаемые апостериорные вероятности успеха. Однако, в функциональном подходе были рычаги защиты от неверных предположений: отклонения от начального распределения и отклонения от найденной зависимости. В стохастическом подходе при выборе максимальной копулы, при отсутствии дополнительных усложнений функции оценки, таких рычагов нет.

4.5 Farli-Gumbel-Morgenstern семейство копул.

Параметрическое семейство Farli-Gumbel-Morgenstern копул с параметром $-1 \leq \eta \leq 1$ задаётся как

$$C(u_1, u_2, \eta) = u_1 u_2 + \eta u_1 u_2 (1 - u_1)(1 - u_2).$$

При этом функция плотности копулы есть

$$c(u_1, u_2, \eta) = 1 + \eta(1 - 2u_1)(1 - 2u_2).$$

Значения параметра $\eta = 1$ и $\eta = -1$ соответствуют случаю наибольшей и наименьшей корреляции между величинами соответственно.

Значение параметра $\eta = 0$ соответствует случаю независимой копулы $I(u_1, u_2) = u_1 u_2$. Нетрудно заметить, что если плотность совместного распределения представляется в виде произведения маргинальных плотностей, то плотность копулы совместного распределения равна 1 всюду на квадрате $[0, 1]^2$. Это означает, что функция распределения копулы есть $I(u_1, u_2) = u_1 u_2$.

Пример копулы на рис. 5.

4.6 Гауссовские копулы.

Гауссовская копула получается из многомерного нормального распределения путем исключения маргинальных распределений. Пусть есть нормальный случайный вектор с ковариационной матрицей $\Sigma \in \mathbb{R}^{d \times d}$. Пусть Φ_Σ — совместная функция распределения, соответствующая распределению с матрицей ковариации Σ и нулевым матожиданием $\mathcal{N}(0, \Sigma)$. Пусть Φ — функция распределения стандартной нормальной случайной величины. Тогда соответствующая вектору копула представляется как:

$$C_\Sigma^{Gauss}(u) = \Phi_\Sigma(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)).$$

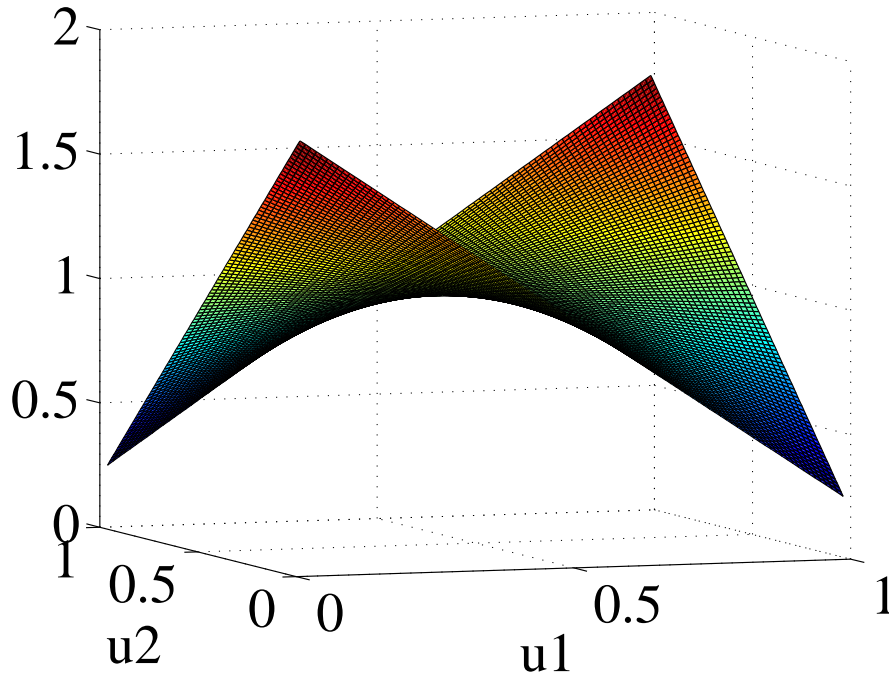


Рис. 5: Плотность распределения копулы Farli-Gumbel-Morgenstern с параметром $\eta = 0.7$.

Плотность распределения копулы имеет вид

$$c_{\Sigma}^{Gauss}(u) = \frac{1}{\sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_d) \end{pmatrix}^T (\Sigma^{-1} - I) \begin{pmatrix} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_d) \end{pmatrix} \right\},$$

где I — единичная матрица.

Пример копулы на рис. 6.

4.7 t копула Стьюдента

Двумерная копула Стьюдента с параметрами ρ и ν задаётся формулой:

$$C_{\rho, \nu}(u_1, u_2) = \int_{-\infty}^{t_{\nu}^{-1}(u_1)} \int_{-\infty}^{t_{\nu}^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \left\{ 1 + \frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{\nu(1-\rho^2)} \right\}^{-(\nu+2)/2} dx_1 dx_2,$$

где $t_{\nu}^{-1}(u)$ — u -квантиль распределения Стьюдента с параметром ν . В общем случае копула Стьюдента задаётся матрицей P и параметром ν . В двумерном случае, соответствующая параметру ρ матрица есть

$$P = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

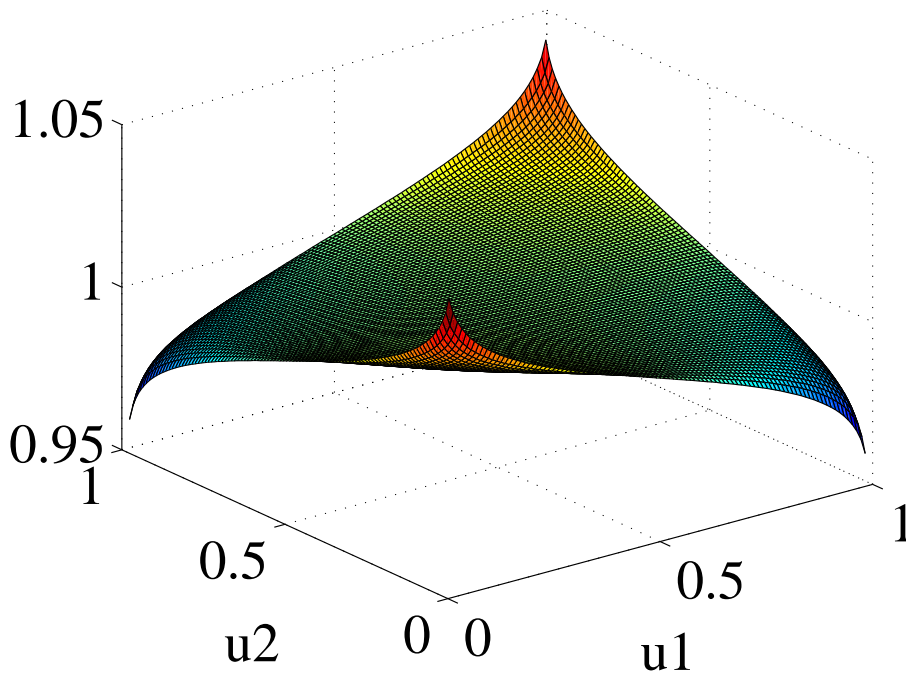


Рис. 6: Плотность распределения Гауссовской копулы с параметром $\rho = 0.008$.

Пример копулы на рис. 7.

4.8 Архимедовы копулы.

Пусть функция $\psi(\cdot)$ удовлетворяет следующим свойствам:

- $\psi(0) = 1$.
- $\lim_{u \rightarrow 0} \psi(u) = \infty$.
- $\psi'(u) < 0$.
- $\psi''(u) > 0$.

Тогда функция

$$C_\psi(u_1, u_2) = \psi^{-1}(\psi(u_1) + \psi(u_2)),$$

является копулой. Такая копула называется архимедовой копулой, а функция ψ — её генератором. Ниже мы рассмотрим несколько наиболее известных семейств архимедовых копул.

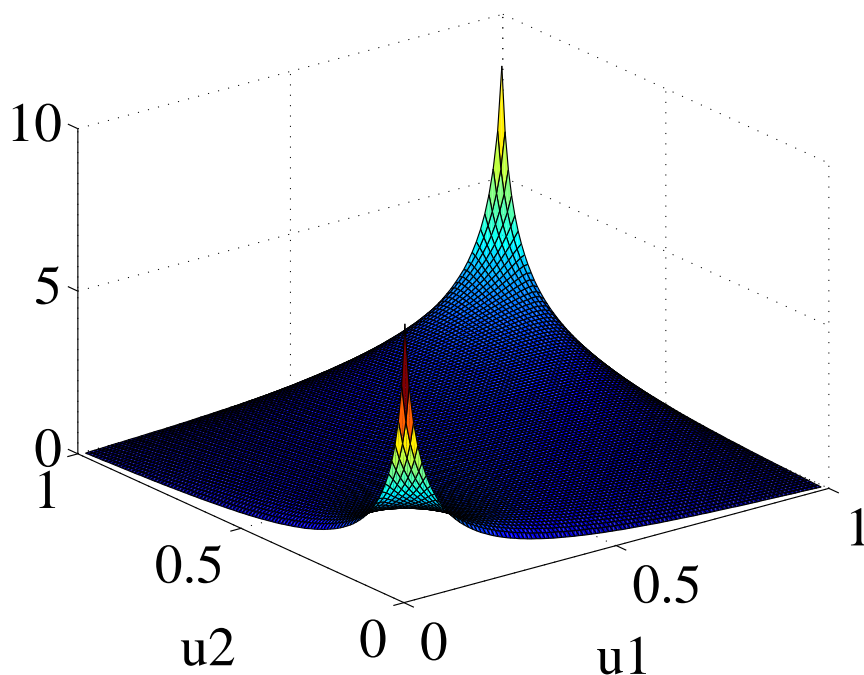


Рис. 7: Плотность распределения t копулы с параметрами $\rho = 0.52$, $\nu = 22$.

4.9 Семейство Clayton копул.

Копула семейства Clayton с параметром $\zeta \in (0, \infty)$ задаётся функцией-генератором:

$$\psi(u, \zeta) = (1 + \zeta u)^{-1/\zeta}, \quad \psi^{-1}(u, \zeta) = \frac{1}{\zeta}(u^{-\zeta} - 1).$$

Пример копулы на рис. 8.

4.10 Семейство Frank копул.

Копула семейства Frank с параметром $\zeta \in (0, \infty)$ задаётся функцией-генератором:

$$\psi(u, \zeta) = -\frac{1}{\zeta} \ln(1 - (1 - \exp(-\zeta)) \exp(-u)), \quad \psi^{-1}(u, \zeta) = -\ln\left(\frac{\exp(-\zeta u) - 1}{\exp(-\zeta) - 1}\right).$$

Пример копулы на рис. 10.

4.11 Семейство Gumbel копул.

Копула семейства Gumbel с параметром $\zeta \in [1, \infty)$ задаётся функцией-генератором:

$$\psi(u, \zeta) = \exp(-u^{1/\zeta}), \quad \psi^{-1}(u, \zeta) = (-\ln u)^\zeta.$$

Пример копулы на рис. ??.

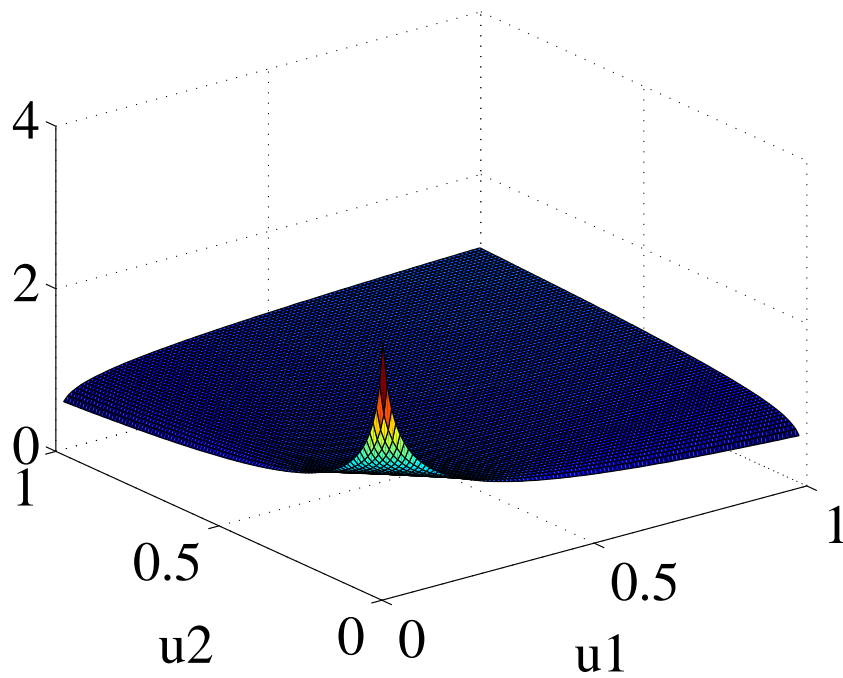


Рис. 8: Плотность распределения Clayton копулы с параметром $\zeta = 0.135$.

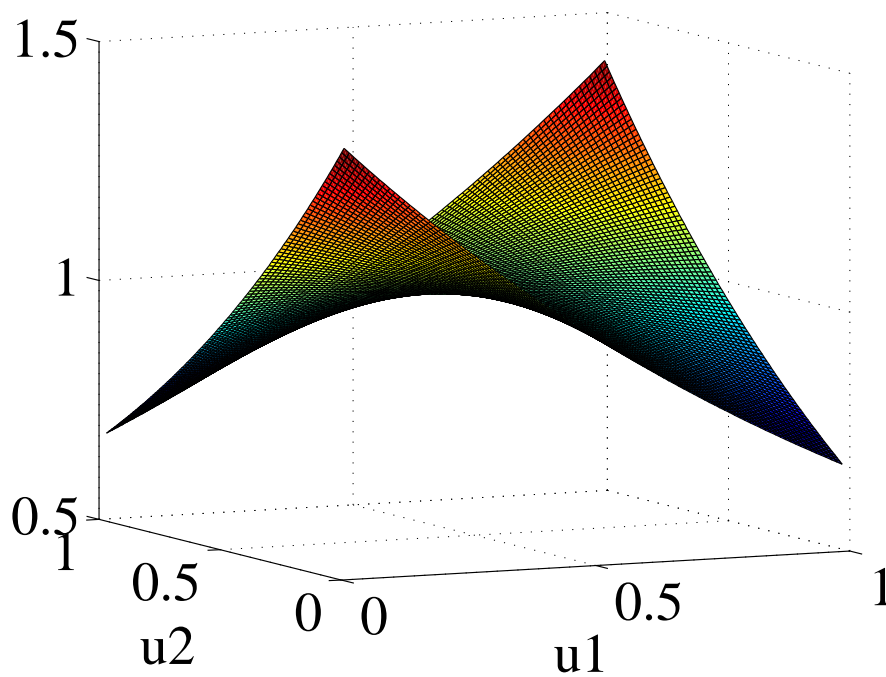


Рис. 9: Плотность распределения Frank копулы с параметром $\zeta = 0.75$.

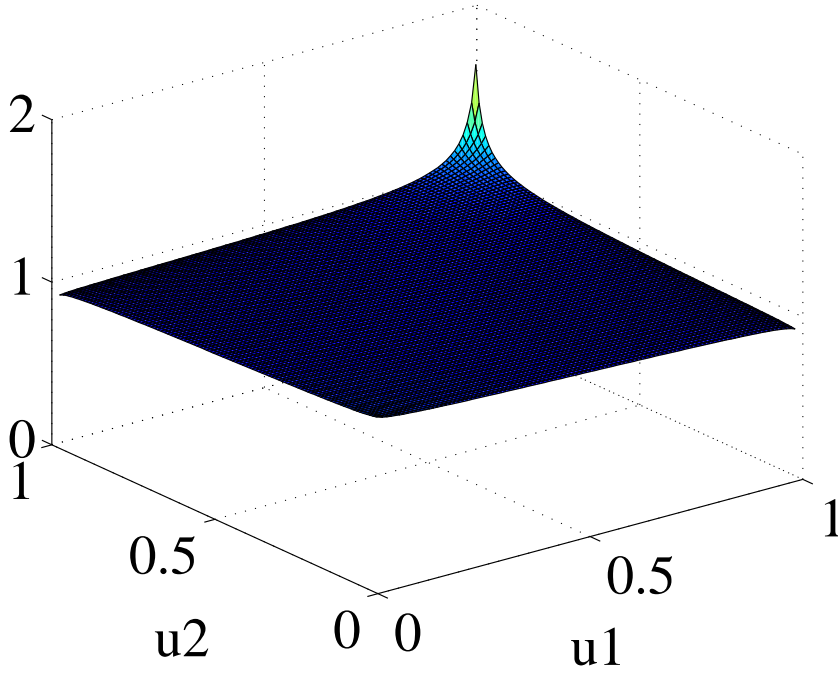


Рис. 10: Плотность распределения Frank копулы с параметром $\zeta = 1.014$.

4.12 Вычисление апостериорного распределения.

Как говорилось выше, функция оценки задаётся формулой

$$f(X, \xi, \zeta, \eta) = \frac{\int_0^1 \int_0^1 pv(p, r) dC(H_1(p, \xi), H_2(r, \zeta), \eta)}{\int_0^1 \int_0^1 v(p, r) dC(H_1(p, \xi), H_2(r, \zeta), \eta)}$$

где $v(p, r) = p^{k_i}(1-p)^{n_i-k_i}r^{l_i}(1-r)^{m_i-l_i}$.

Таким образом, функция оценки есть отношение матожиданий двух случайных величин: $pv(p, r)$ и $v(p, r)$. Рассмотрим следующий метод Монте Карло вычисления матожидания функции $g(x_1, x_2)$ по совместному распределению, заданному копулой $C(u_1, u_2)$ и маргинальными распределениями $H_1(x_1)$ и $H_2(x_2)$. Подобная процедура описана в [10].

1. Сгенерировать множество пар $\{u_1^k, u_2^k\}_{k=1}^N$ из распределения $C(u_1, u_2)$.
2. Получить множество пар $\{x_1^k, x_2^k\}$ преобразованиями

$$x_1^k = H_1^{-1}(u_1^k), \quad x_2^k = H_2^{-1}(u_2^k).$$

3. Вычислить значения $g^k = g(x_1^k, x_2^k)$.

4. Получить оценку искомого матожидания

$$Eg(\widehat{x_1, x_2}) = \frac{1}{N} \sum_{k=1}^N g^k.$$

5 Вычислительный эксперимент.

5.1 ROC-кривые.

В дополнение к значениям целевого функционала качества для функций оценки были построены ROC-кривые и вычислены площади AUC под ними. ROC-кривая задаётся как однопараметрическая кривая с параметром t . Пусть выбрана некоторая функция оценки $f(X)$ и значение $t \in [0, 1]$. Рассмотрим таблицу 5.1,

	$a_i = 1$	$a_i = 0$
$f(X_i) > t$	b_{11}	b_{10}
$f(X_i) \leq t$	b_{01}	b_{00}

в которой $b_{j_1 j_2}$ есть количество соответствующих объектов, т.е.

$$b_{j_1 j_2}(t) = \#[[I(f(X_i) > t) = j_1] \& [a_i = j_2]], \quad I(true) = 1, \quad I(false) = 0.$$

Тогда значения *False-Positive rate* и *True-Positive rate* определяются как

$$FPR(t) = \frac{b_{10}(t)}{b_{10}(t) + b_{00}(t)},$$

$$TPR(t) = \frac{b_{11}(t)}{b_{11}(t) + b_{01}(t)}.$$

ROC-кривая есть множество точек $(FPR(t), TPR(t))$ для $t \in [0, 1]$.

5.2 Функциональный подход.

Оптимизация проводилась методом градиентного спуска.

Выборка $\{x_1, \dots, x_M\} = \{(X_i, a_i)\}_{i=1}^M$ была преобразована в выборку $\{(X_j, c_j, s_j)\}_{j=1}^N$, где совокупность различных X_i совпадает с множеством $\{X_j\}_{j=1}^N$, $c_j = \sum_{i: X_i = X_j} a_i$ — количество успехов в дополнительном испытании среди объектов с

$X_i = (n_i, k_i, m_i, l_i) = X_j$, $s_j = \sum_{i: X_i = X_j} (1 - a_i)$ — количество неуспехов в дополнительном испытании среди объектов с $X_i = (n_i, k_i, m_i, l_i) = X_j$. Функционал качества в терминах новой выборки записывается как

$$\mathcal{L}(f) = \sum_j (c_j \ln f(X_j) + s_j \ln(1 - f(X_j))).$$

Соответственно, все формулы для производных изменяются следующим образом: суммирование ведётся по последовательности X_j , а не X_i , a_i заменяется на c_j , $(1 - a_i)$ заменяется на s_j .

Для нахождения оптимальных значений параметров использовался алгоритм градиентного спуска:

1. Вход: выборка $\mathbb{X} = \{(X_j, c_j, s_j)\}$, вектор начальных значений параметров $\theta \in \mathbb{R}^d$, параметр сжатия ϵ , критерий останковки δ .
2. Шаг 1: Вычислить $Q = \mathcal{L}(f(X, \theta))$.
3. Шаг 2: Вычислить производные $\frac{\partial \mathcal{L}}{\partial \theta_i}$ для $i = 1 \dots d$.
4. Шаг 3: Обновить значения $\theta_i = \theta_i + \epsilon \frac{\partial \mathcal{L}}{\partial \theta_i}$ для $i = 1 \dots d$.
5. Шаг 4: Вычислить новое значение функционала $R = \mathcal{L}(f(X, \theta))$.
6. Шаг 5: Если $|R - Q| > \delta$, присвоить $Q = R$ и вернуться на шаг 2.
7. Выход: θ, R .

В таблице 5.2 показаны значения целевого функционала \mathcal{L} и площади AUC под ROC-кривой для различных функций оценки. При этом использованы обозначения *lin*, *log*, *rat* для линейной, логарифмической и рациональной функции веса; обозначения *lin*, *sig* использованы для обозначения линейной и сигмоидальной функции зависимости. Так, например, оптимальная оценка из класса оценок функционального подхода с рациональной функцией веса и сигмоидальной функцией зависимости обозначается *rat sig*. Обозначение f^* используется для оценки верхней границы в рассматриваемом классе функций. Обозначение g^* используется для оценки верхней границы в редуцированном классе функций, зависящих только от аргументов n и k .

Графики ROC-кривых и средних значений функционала правдоподобия можно увидеть на рис. 11–22, 37–44.

5.3 Стохастический подход.

Для нахождения оптимальных значений параметров в стохастическом подходе, задача разбивается на три части (такой подход описан в [11]):

1. Оценка вектора параметров ξ маргинального Бета распределения $H_1(p, \xi)$ вероятности успеха в первой последовательности испытаний.

f	$AUC(f)$	$\mathcal{L}(f)$
f^*	0.714	-0.306
$lin\ lin$	0.681	-0.316
$lin\ sig$	0.692	-0.313
$log\ lin$	0.688	-0.315
$log\ sig$	0.696	-0.312
$rat\ lin$	0.690	-0.314
$rat\ sig$	0.700	-0.311
g^*	0.682	-0.316

Таблица 1: Значения $AUC(f)$ площади под ROC-кривой и значения целевого функционала $\mathcal{L}(f)$ для различных оценок функционального подхода.

2. Оценка вектора параметров ζ маргинального Бета распределения $H_2(r, \zeta)$ вероятности успеха во второй последовательности испытаний.
3. Оценка вектора параметров η параметрического семейства $C(u_1, u_2, \eta)$ копул.

Для получения оценки $\hat{\xi}$ рассматривается семейство функций $f(X, \xi) = \frac{k+\xi_0}{n+\xi_1}$ и методом градиентного спуска решается задача $\max_{\xi} \mathcal{L}(f(X, \xi))$.

Для получения оценки $\hat{\zeta}$ рассматривается смежная задача оценивания вероятности успеха, но во второй последовательности испытаний. Для этого исходная выборка пар (X_i, a_i) преобразуется в выборку пар (Y_i, b_i) , где $Y_i = (n_i, k_i, m_i - 1, l_i - b_i)$, b_i — успех дополнительного испытания во второй последовательности испытаний и $b_i = 1$ с вероятностью $\frac{l_i}{m_i}$ и $b_i = 0$ с вероятностью $1 - \frac{l_i}{m_i}$. Далее, рассматривается семейство функций $\tilde{f}(Y, \zeta) = \frac{l+\zeta_0}{m+\zeta_1}$ и методом градиентного спуска решается задача $\max_{\zeta} \tilde{\mathcal{L}}(\tilde{f}(X, \zeta))$.

Для оценки $\hat{\eta}$ сначала получаем выборку оценок

$$(\hat{p}, \hat{r}) : \hat{p}_i = f(X_i, \hat{\xi}), \hat{r}_i = \tilde{f}(X_i, \hat{\zeta}).$$

Множество пар (\hat{p}_i, \hat{r}_i) преобразуется в выборку из распределения копулы:

$$(u_1, u_2) : u_1 = H_1^{-1}(\hat{p}, \hat{\xi}), u_2 = H_2^{-1}(\hat{r}, \hat{\zeta}).$$

Далее, методом *Inference Functions for Margins (IFM)* [12] по выборке пар (u_1^i, u_2^i) получаем оценку $\hat{\eta}$ параметра копулы из выбранного семейства $C(u_1, u_2, \eta)$ копул.

f	$AUC(f)$	$\mathcal{L}(f)$
f^*	0.714	-0.306
max	0.674	-0.317
$F - G - M$	0.682	-0.315
$Clayton$	0.700	-0.311
$Frank$	0.694	-0.313
$Gumbel$	0.693	-0.313
$Gauss$	0.698	-0.312
$Student$	0.702	-0.310
g^*	0.682	-0.316

Таблица 2: Значения $AUC(f)$ площади под ROC-кривой и значения целевого функционала $\mathcal{L}(f)$ для различных оценок стохастического подхода.

После того, как параметры ξ , ζ , η выбраны, как описывалось выше, вычисляем оценку как матожидание по апостериорному распределению методом Монте Карло.

В таблице 5.3 показаны значения целевого функционала \mathcal{L} и площади AUC под ROC-кривой для различных функций оценки. При этом использованы обозначения max , $F - G - M$, $Clayton$, $Frank$, $Gumbel$, $Gauss$, $Student$ для максимальной копулы M и семейств копул *Farli-Gumbel-Morgenstern*, *Clayton*, *Frank*, *Gumbel*, Гауссовских копул, t копул Стьюдента. Обозначение f^* используется для оценки верхней границы в рассматриваемом классе функций. Обозначение g^* используется для оценки верхней границы в редуцированном классе функций, зависящих только от аргументов n и k .

Графики ROC-кривых и средних значений функционала правдоподобия можно увидеть на рис. 23–36, 45–50.

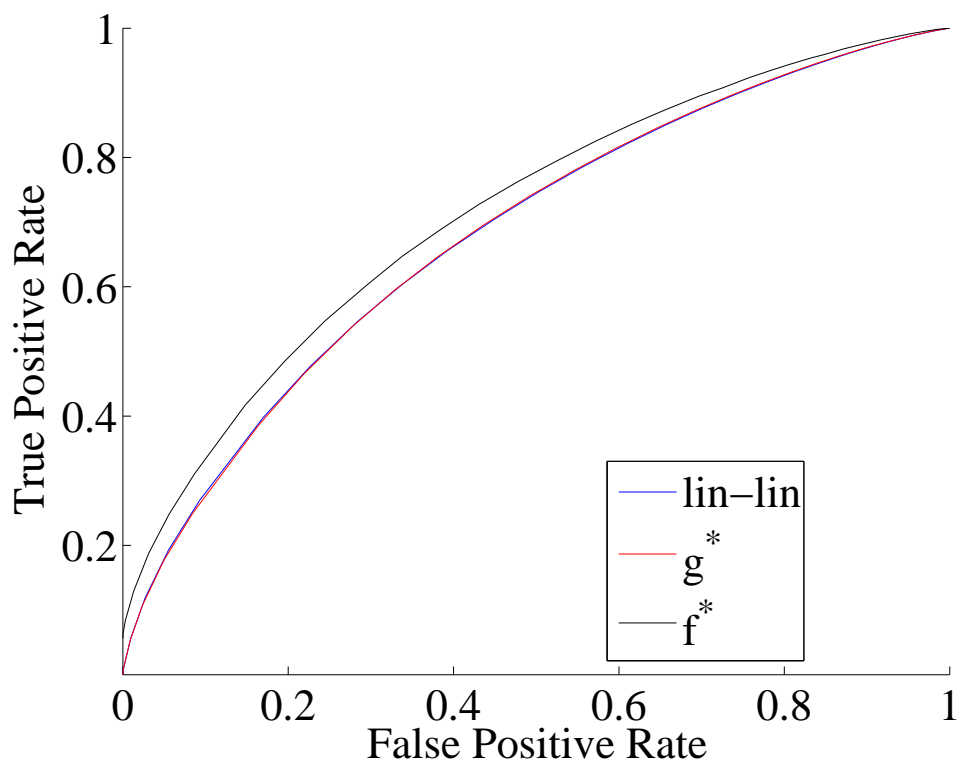


Рис. 11: ROC-кривые для функций оценки f^* , $linlin$, g^* .

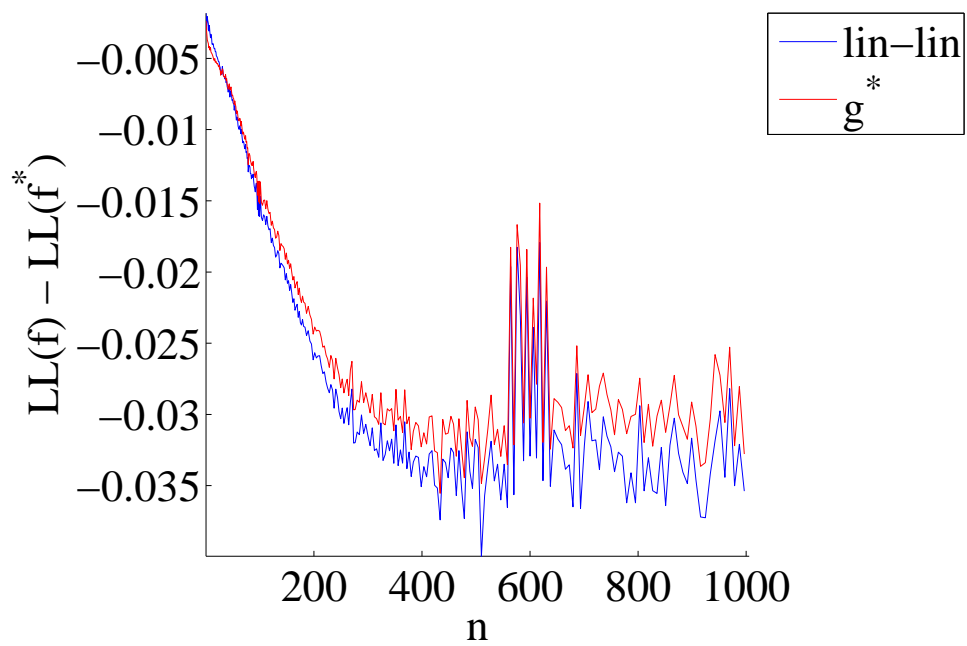


Рис. 12: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки $linlin$, g^* .

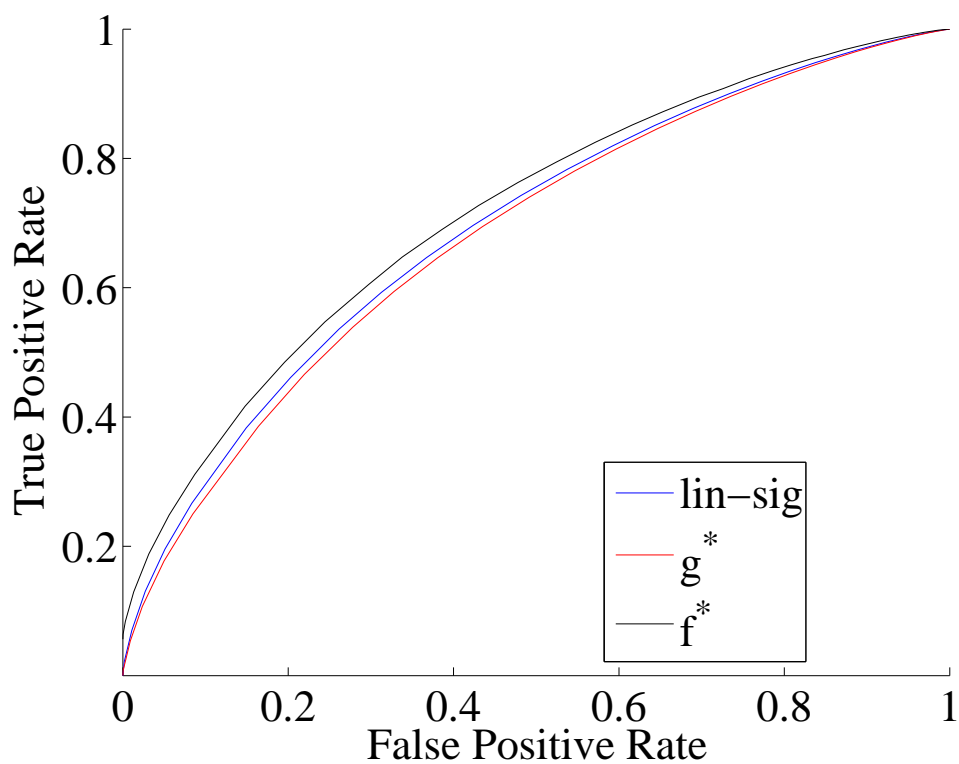


Рис. 13: ROC-кривые для функций оценки f^* , $linsig$, g^* .

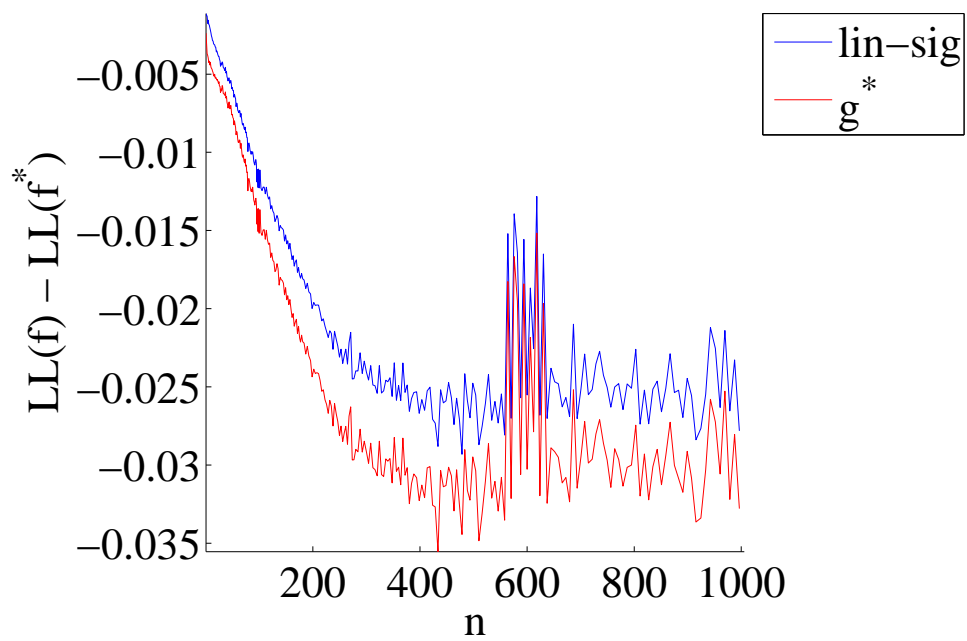


Рис. 14: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки $linsig$, g^* .

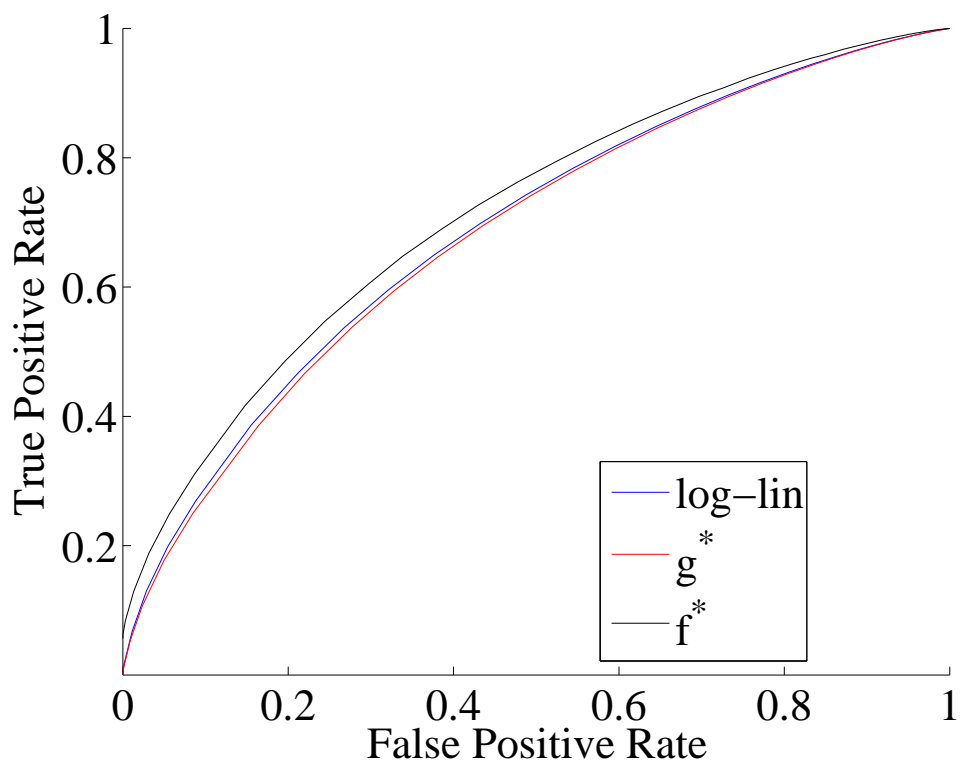


Рис. 15: ROC-кривые для функций оценки f^* , $\log lin$, g^* .

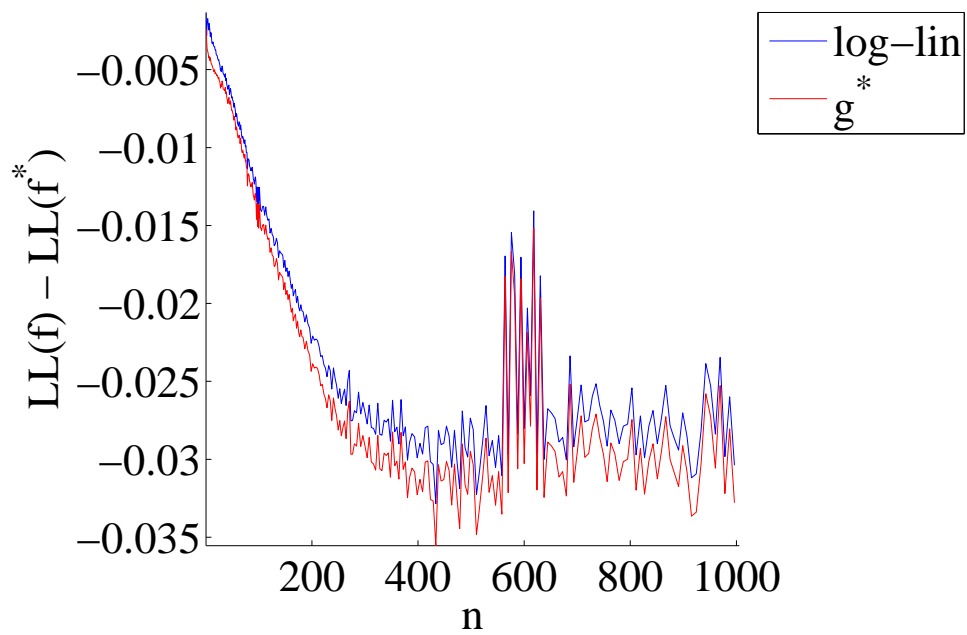


Рис. 16: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки $\log lin$, g^* .

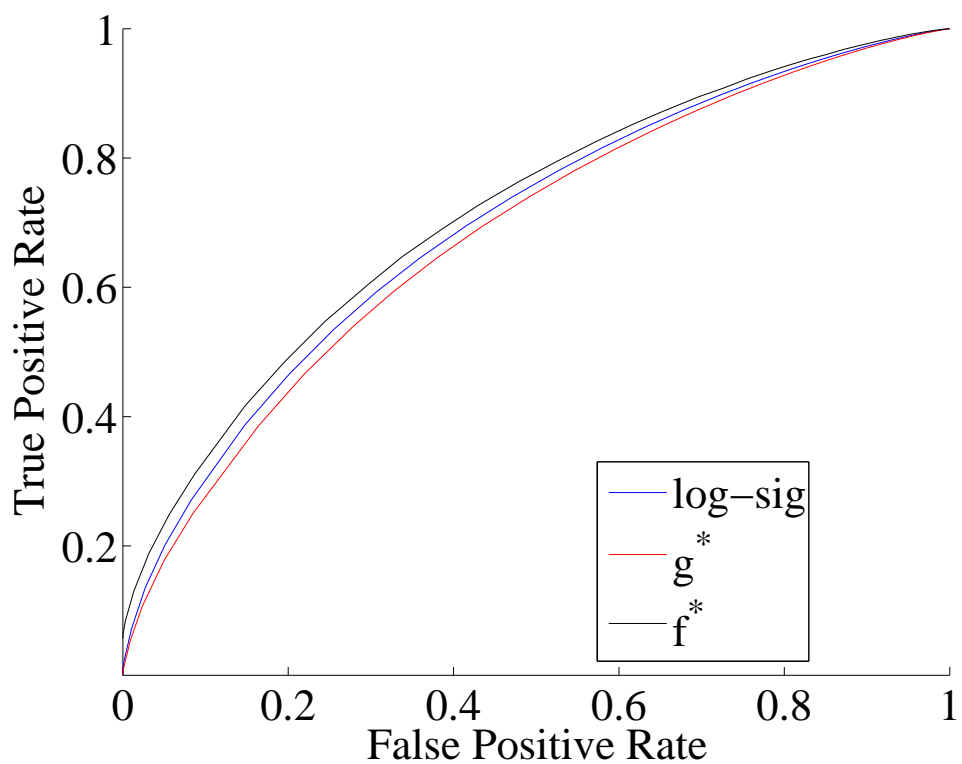


Рис. 17: ROC-кривые для функций оценки f^* , logsig , g^* .

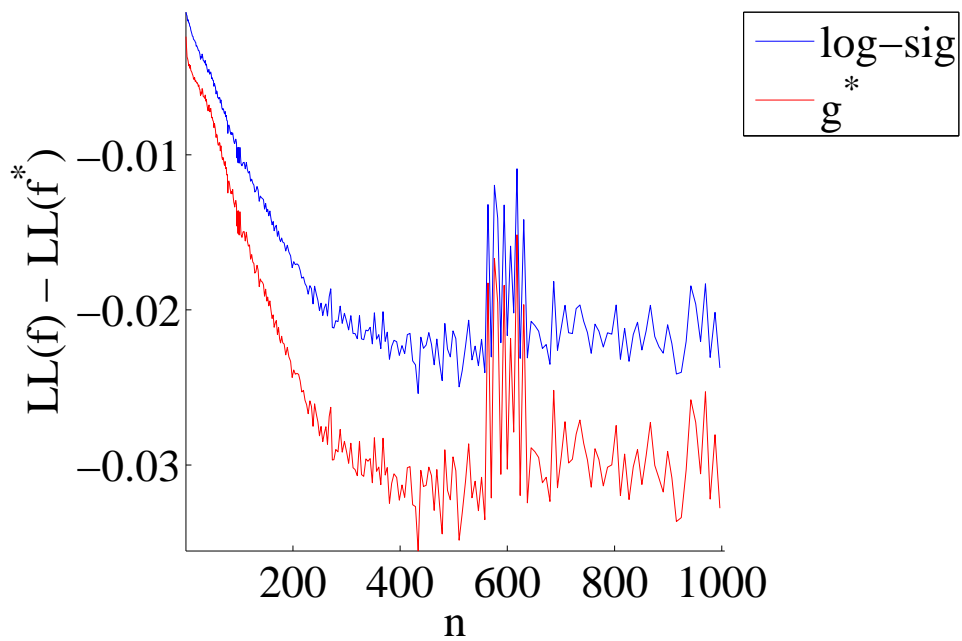


Рис. 18: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки logsig , g^* .

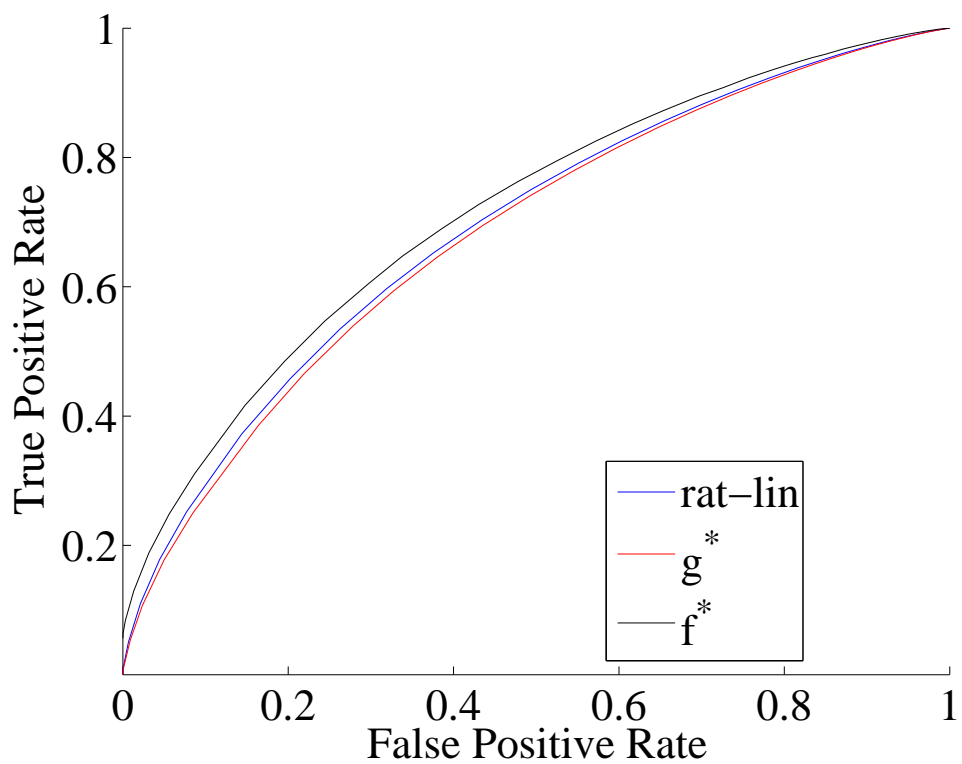


Рис. 19: ROC-кривые для функций оценки f^* , $ratlin$, g^* .

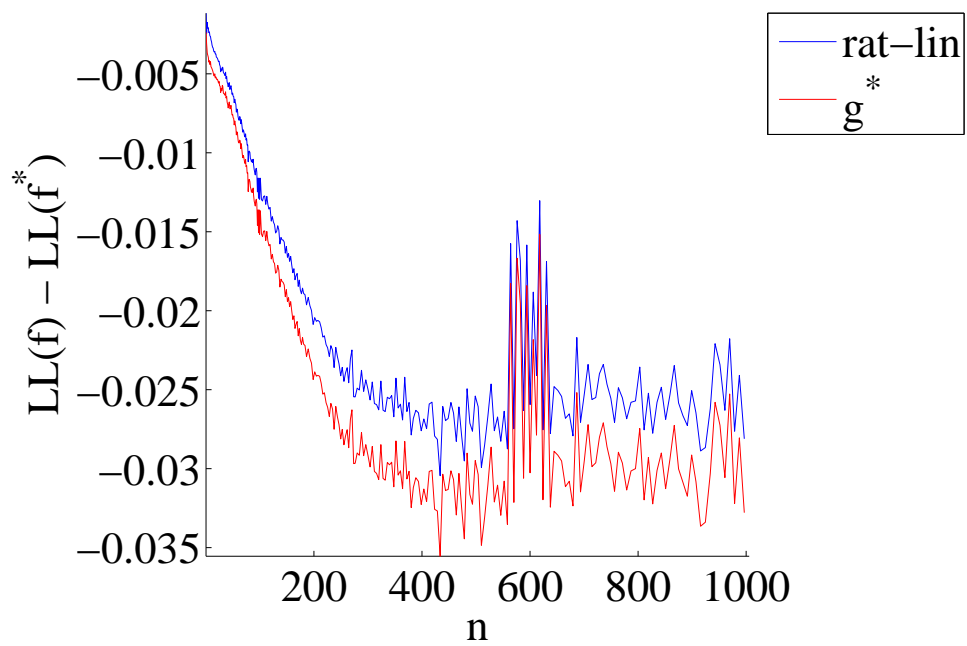


Рис. 20: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки $ratlin$, g^* .

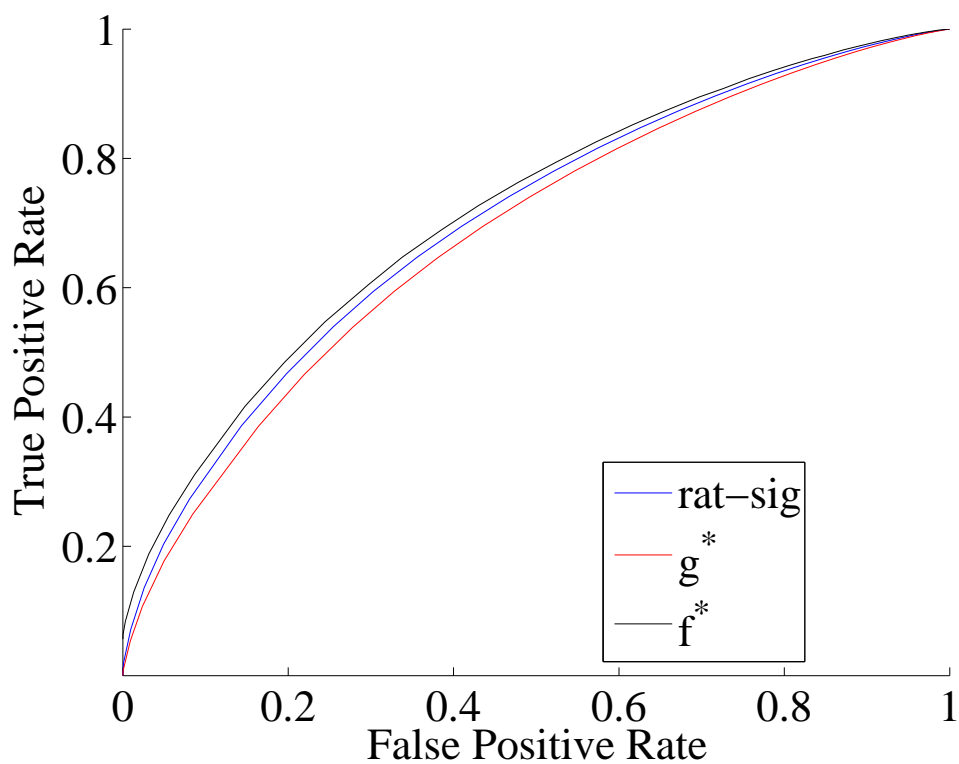


Рис. 21: ROC-кривые для функций оценки f^* , $ratsig$, g^* .

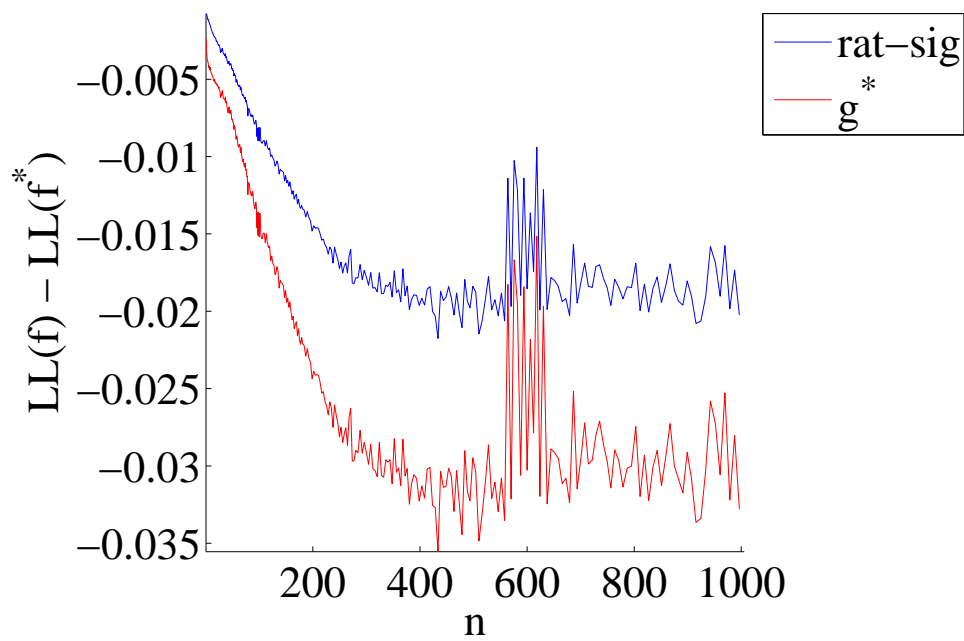


Рис. 22: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки $ratsig$, g^* .

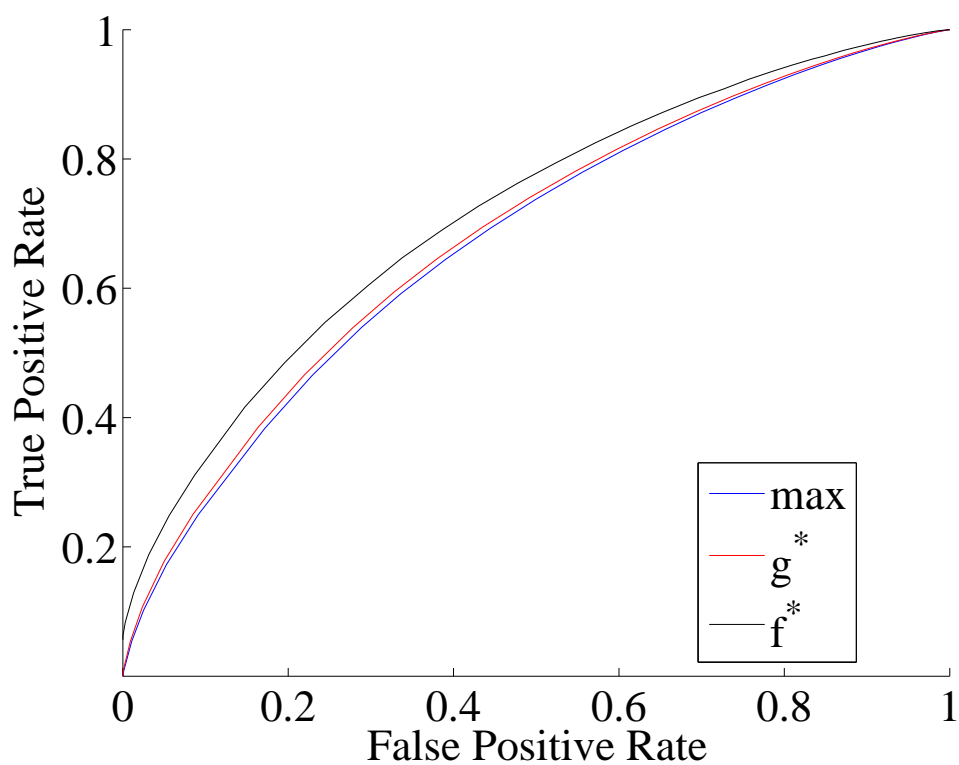


Рис. 23: ROC-кривые для функций оценки f^* , \max , g^* .

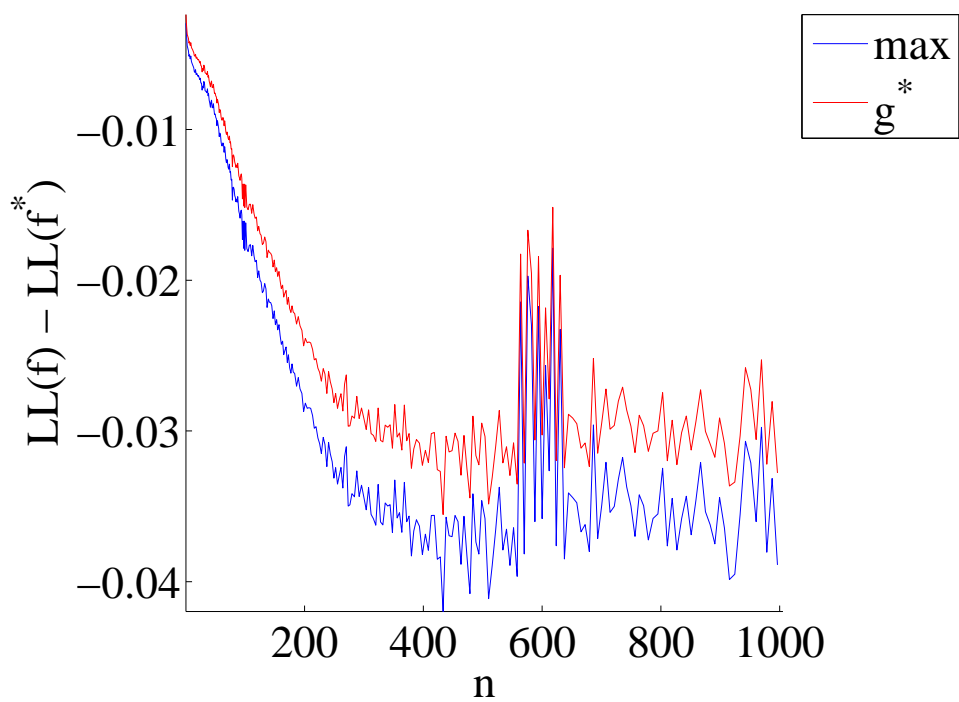


Рис. 24: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки \max , g^* .

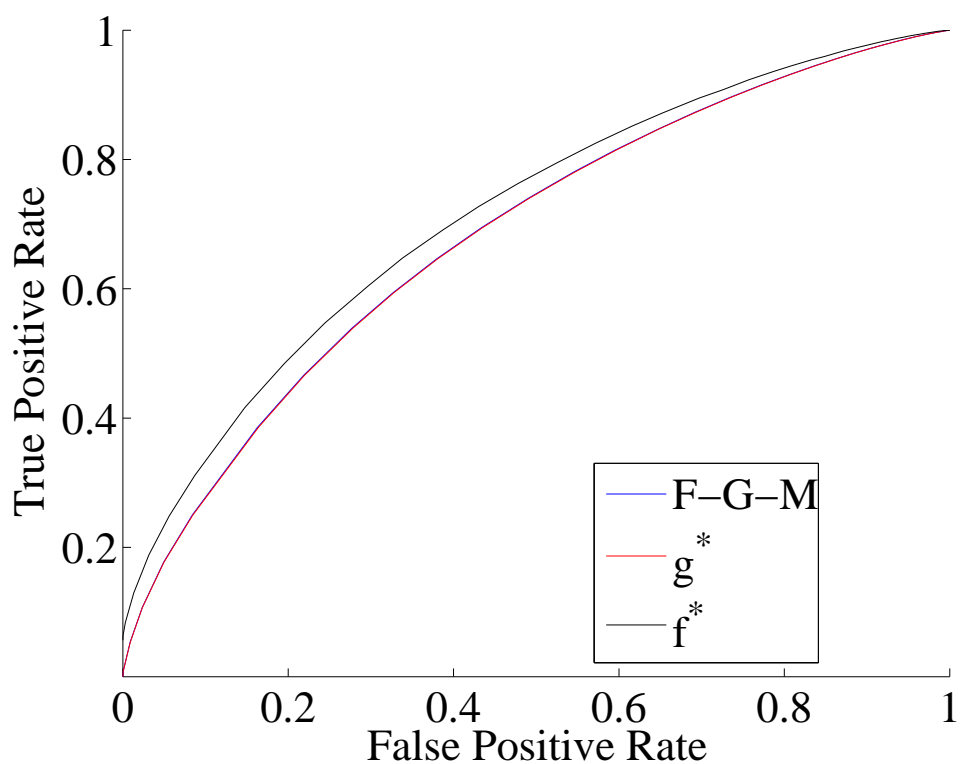


Рис. 25: ROC-кривые для функций оценки f^* , $F - G - M$, g^* .

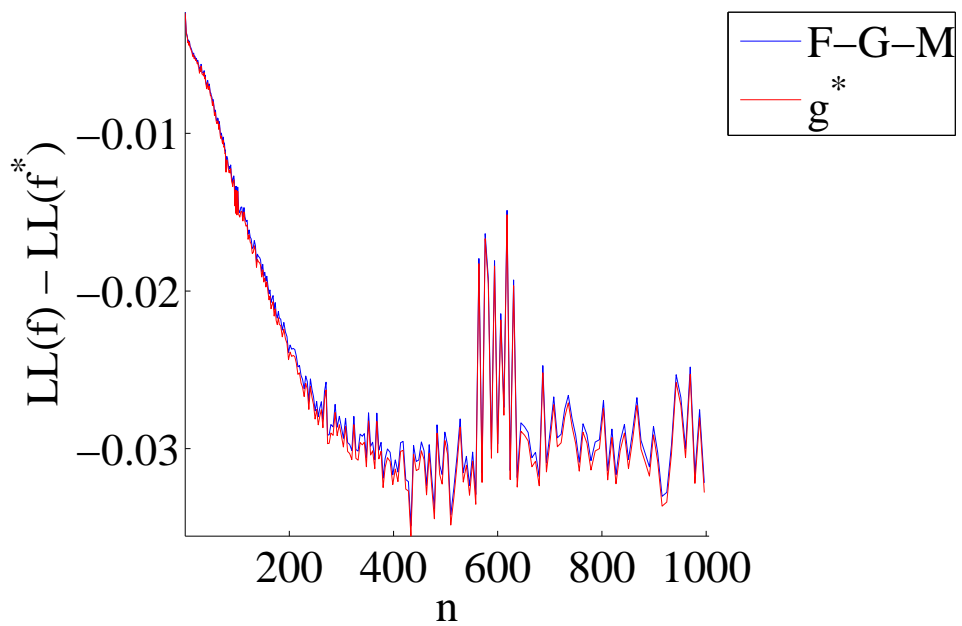


Рис. 26: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки $F - G - M$, g^* .

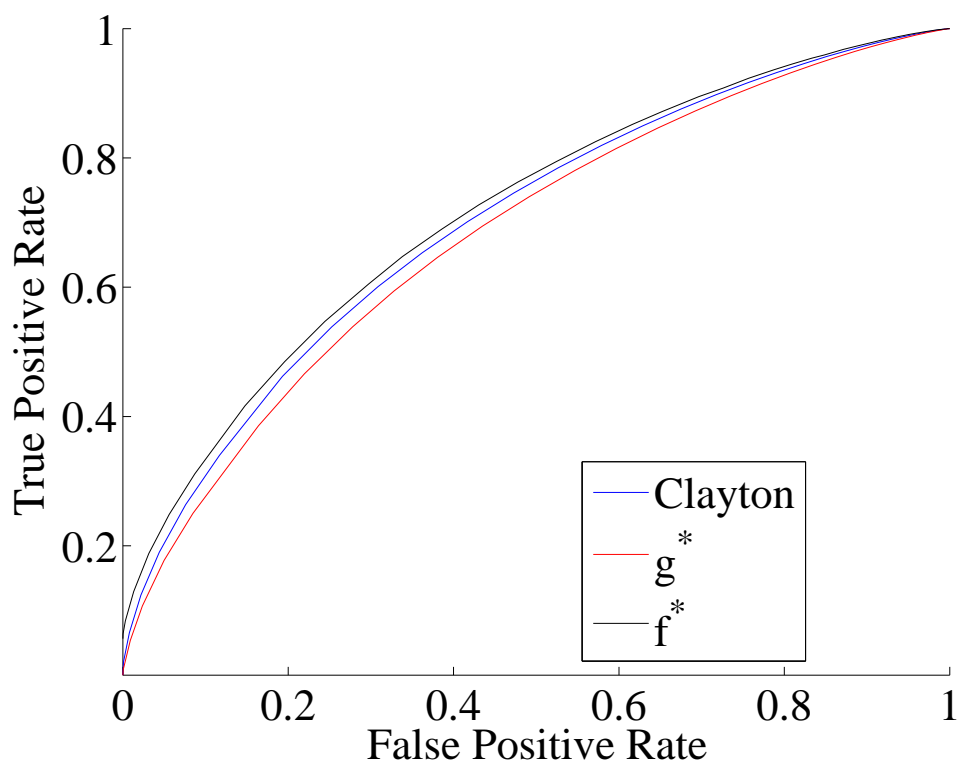


Рис. 27: ROC-кривые для функций оценки f^* , *Clayton*, g^* .

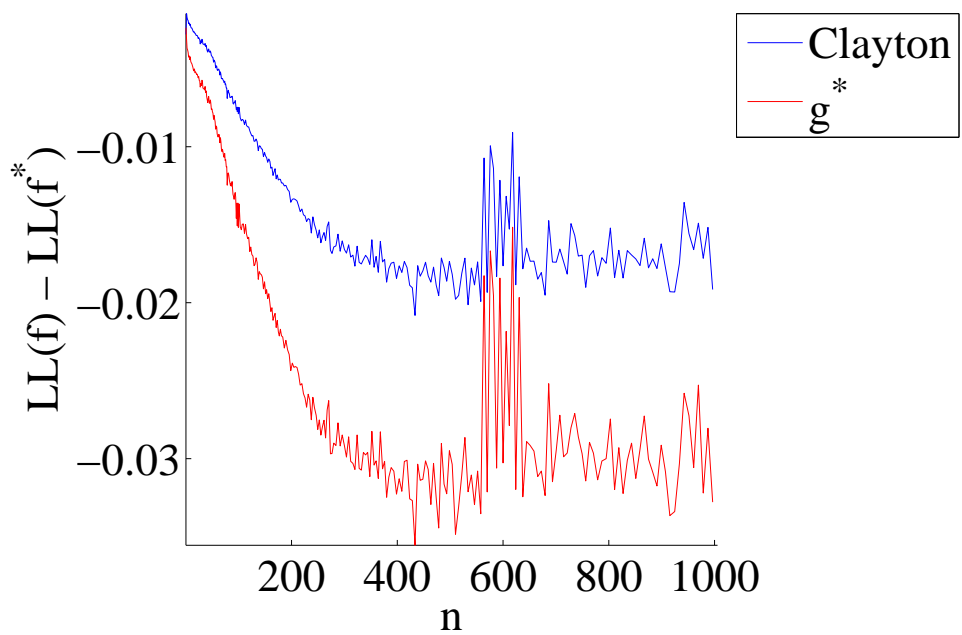


Рис. 28: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки *Clayton*, g^* .

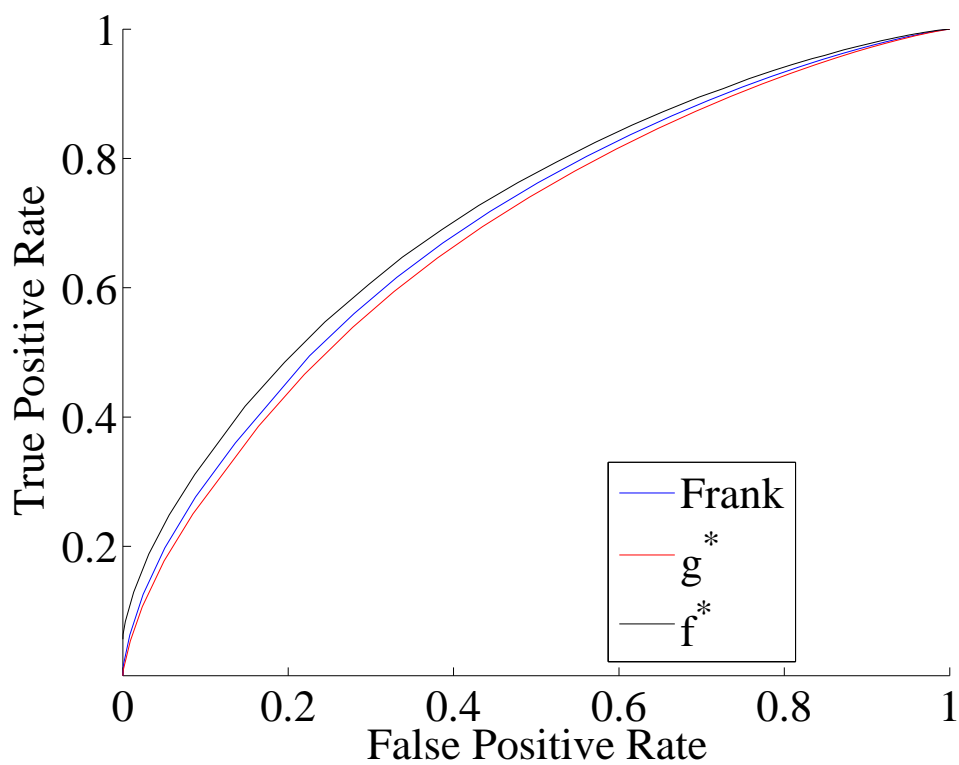


Рис. 29: ROC-кривые для функций оценки f^* , $Frank$, g^* .

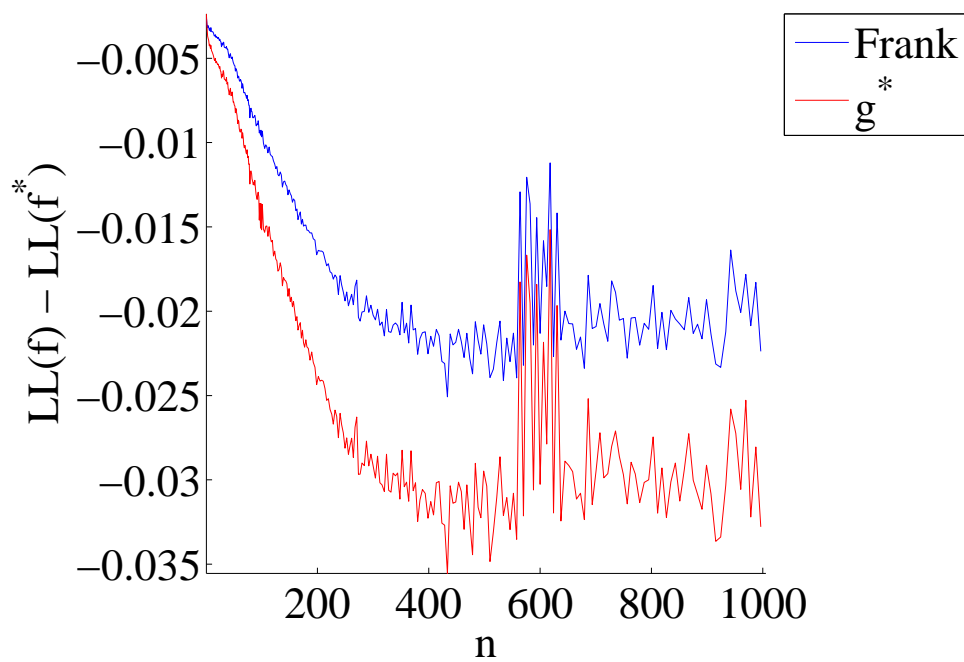


Рис. 30: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки $Frank$, g^* .

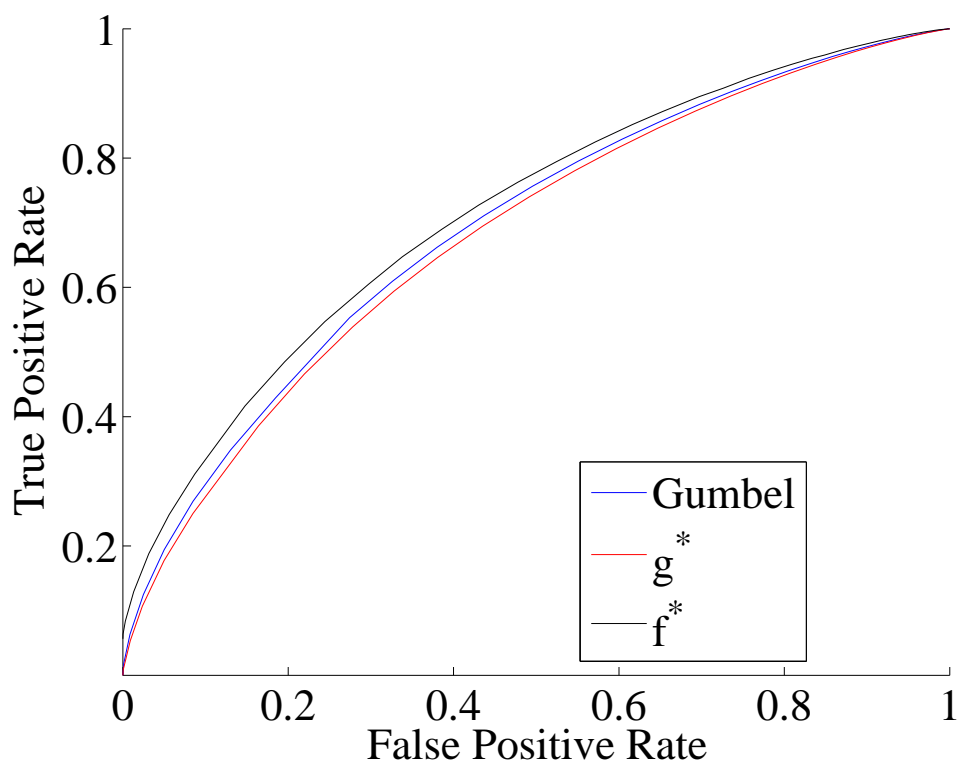


Рис. 31: ROC-кривые для функций оценки f^* , *Gumbel*, g^* .

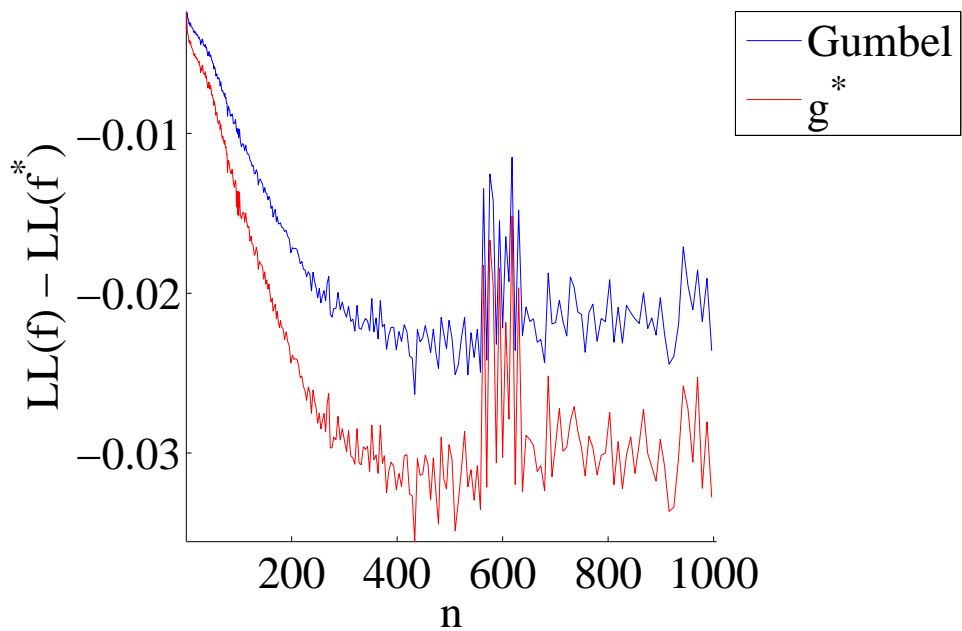


Рис. 32: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки *Gumbel*, g^* .

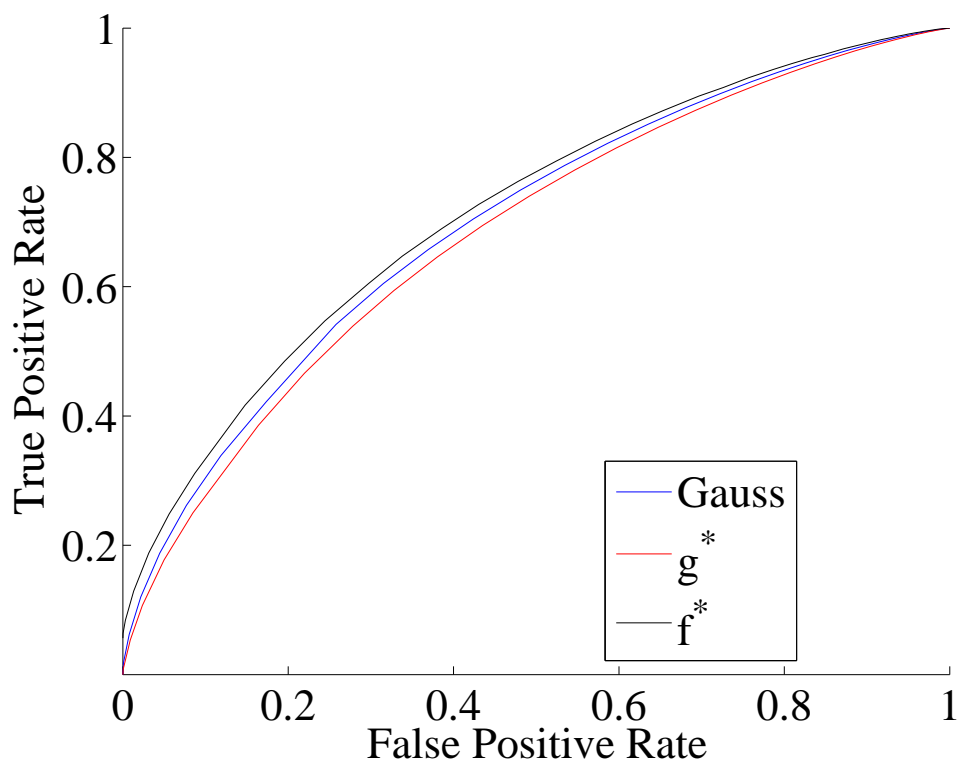


Рис. 33: ROC-кривые для функций оценки f^* , $Gauss$, g^* .

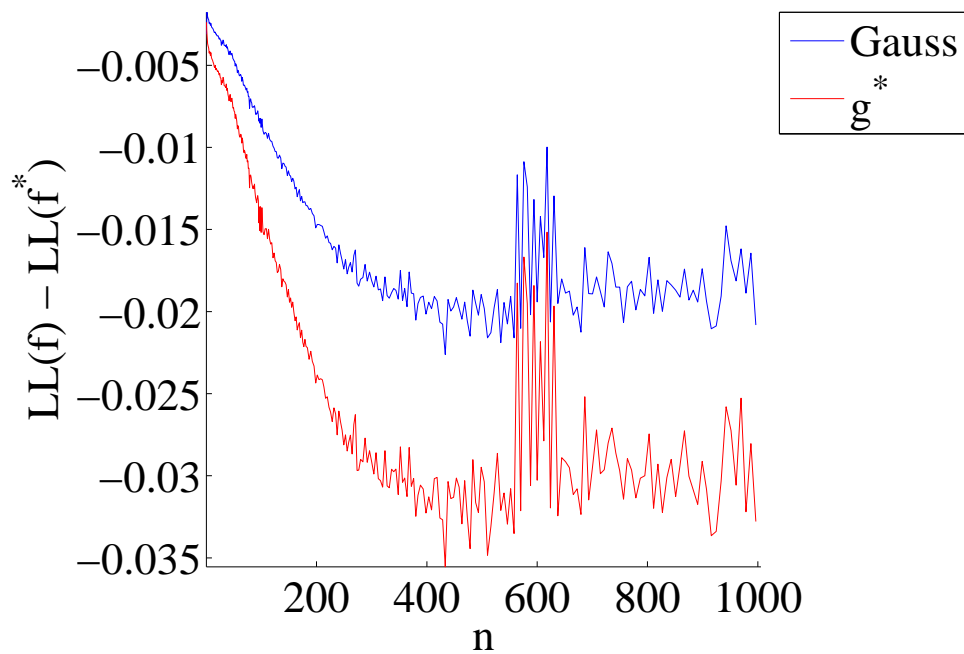


Рис. 34: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки $Gauss$, g^* .

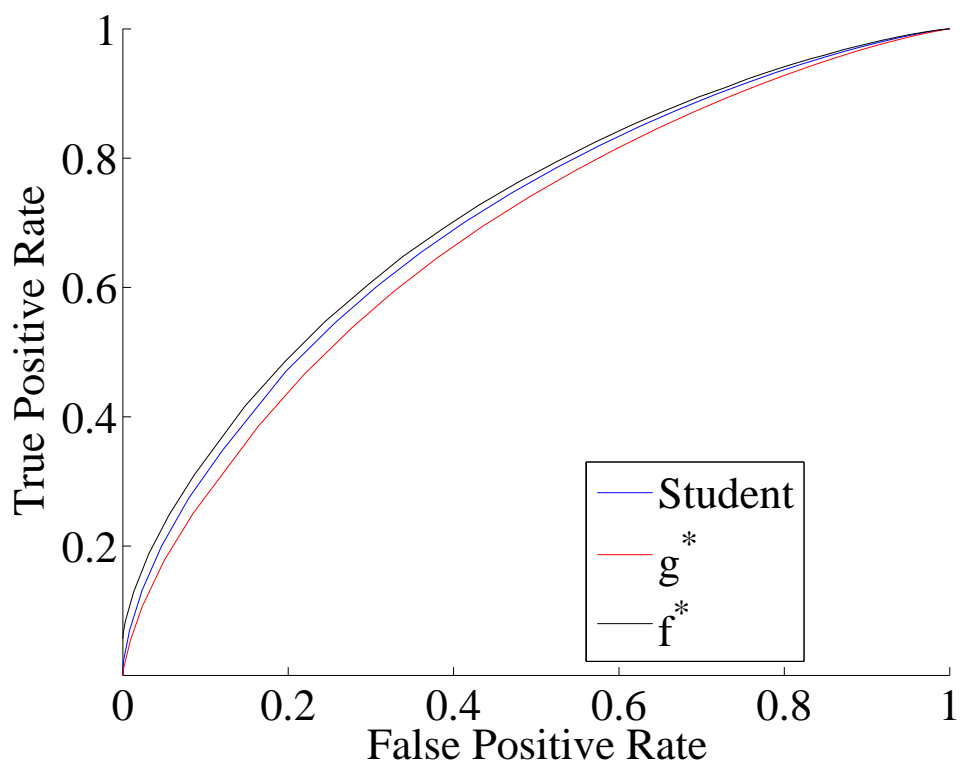


Рис. 35: ROC-кривые для функций оценки f^* , $Student$, g^* .

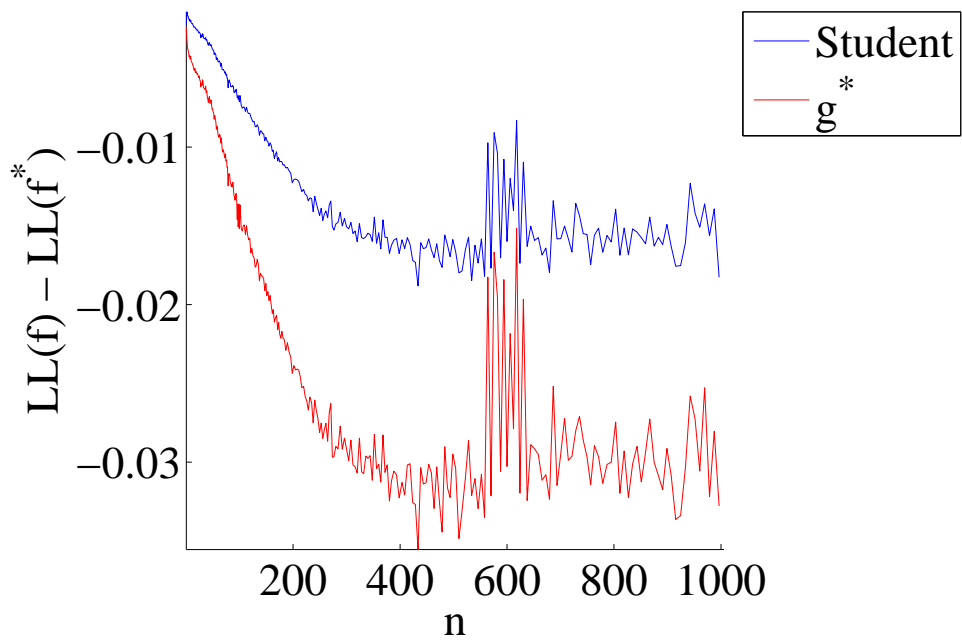


Рис. 36: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки $Student$, g^* .

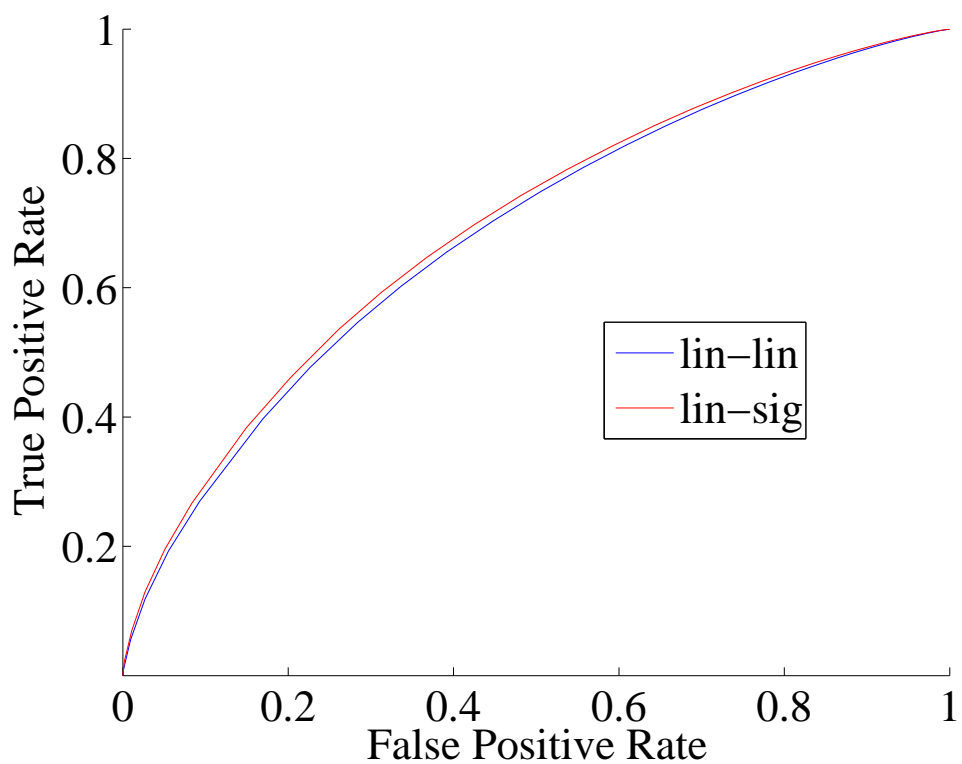


Рис. 37: ROC-кривые для функций оценки $linlin$, $linsig$.

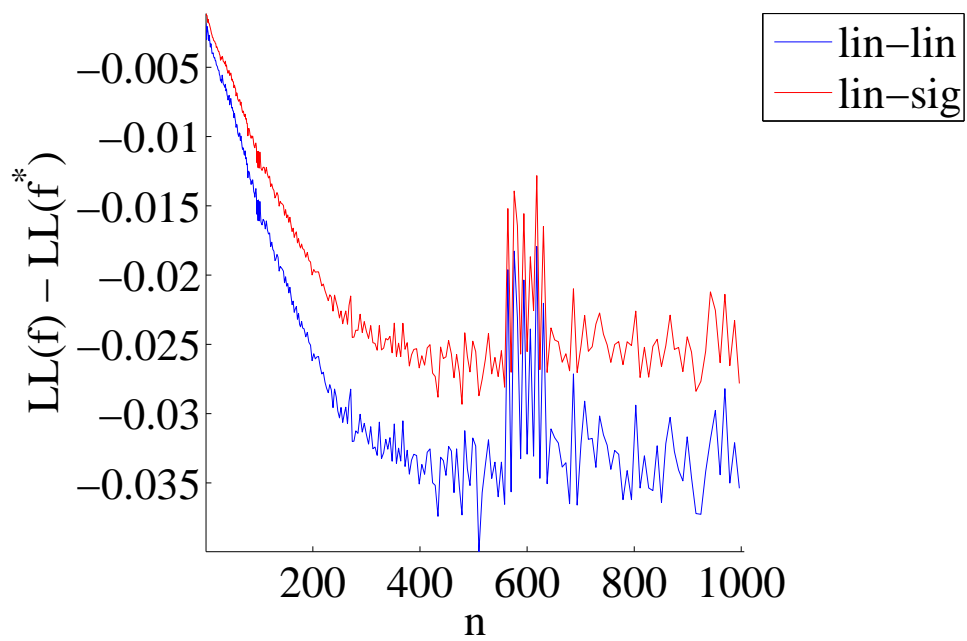


Рис. 38: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки $linlin$, $linsig$.

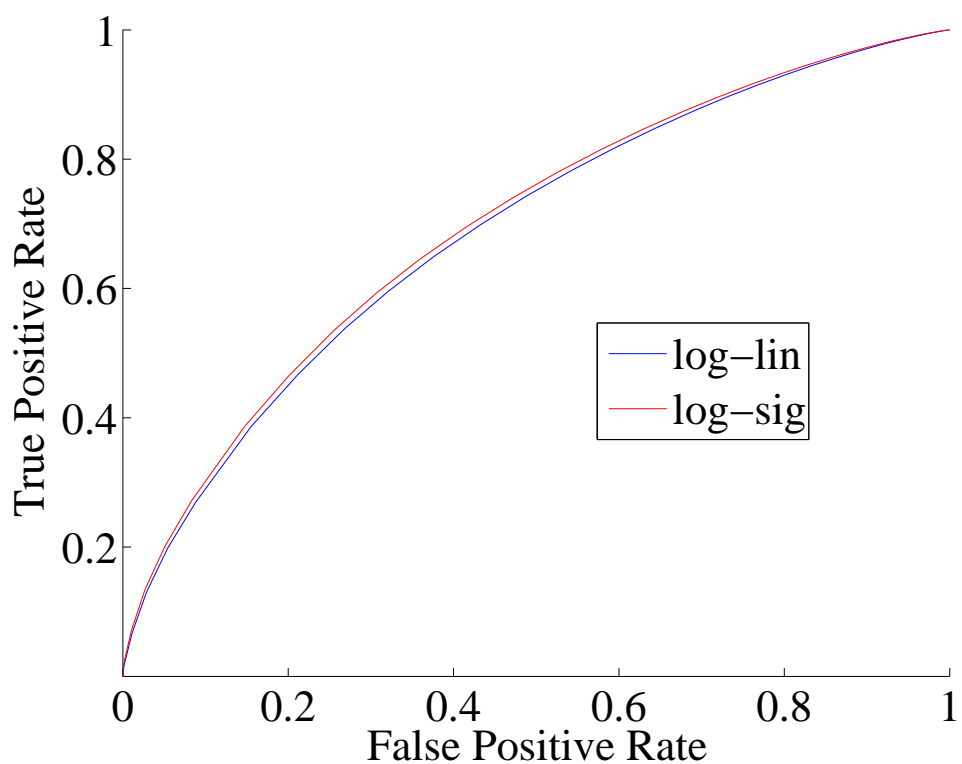


Рис. 39: ROC-кривые для функций оценки loglin , logsig .

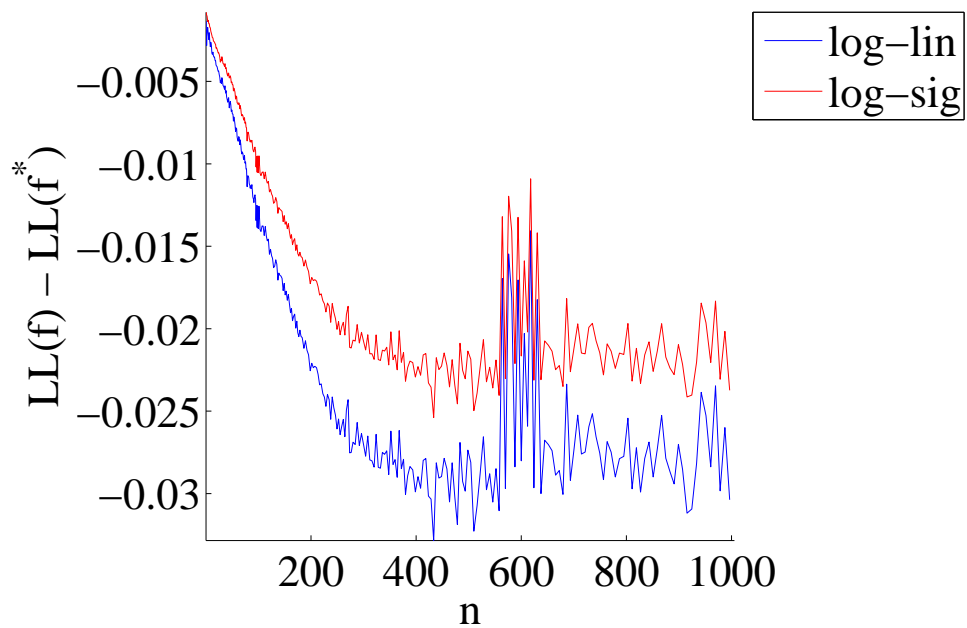


Рис. 40: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки loglin , logsig .

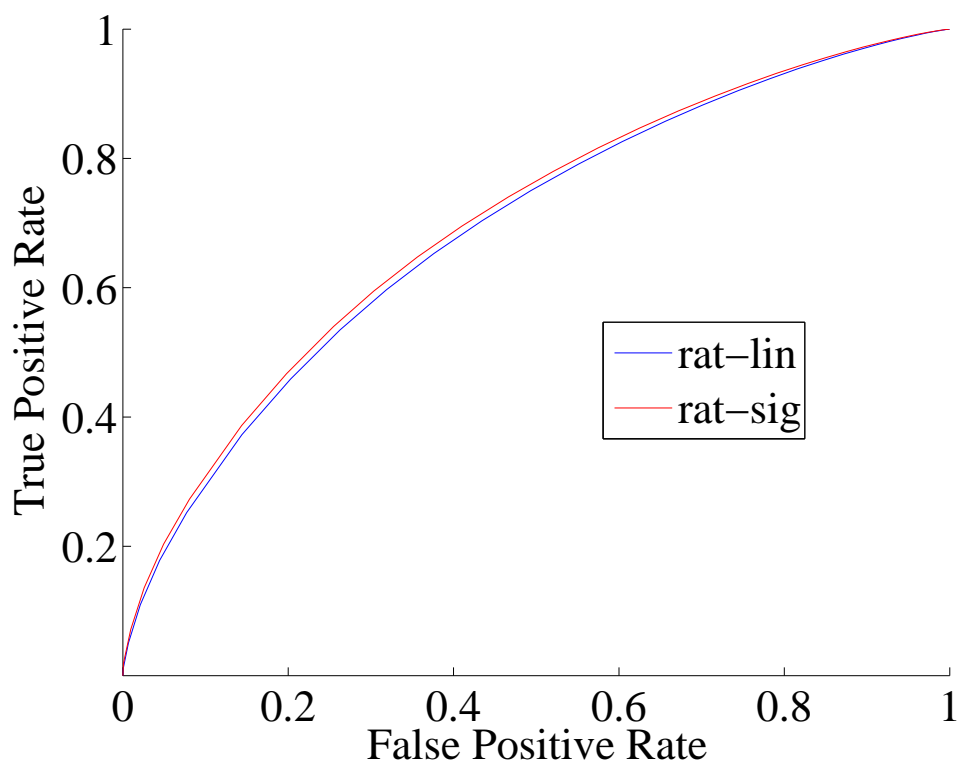


Рис. 41: ROC-кривые для функций оценки $ratlin$, $ratsig$.

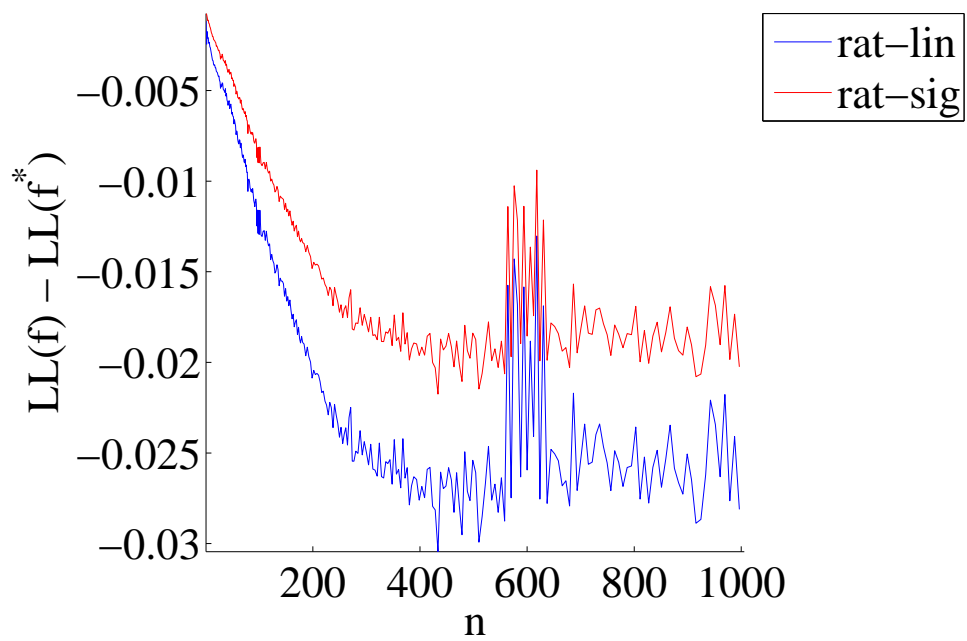


Рис. 42: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки $ratlin$, $ratsig$.

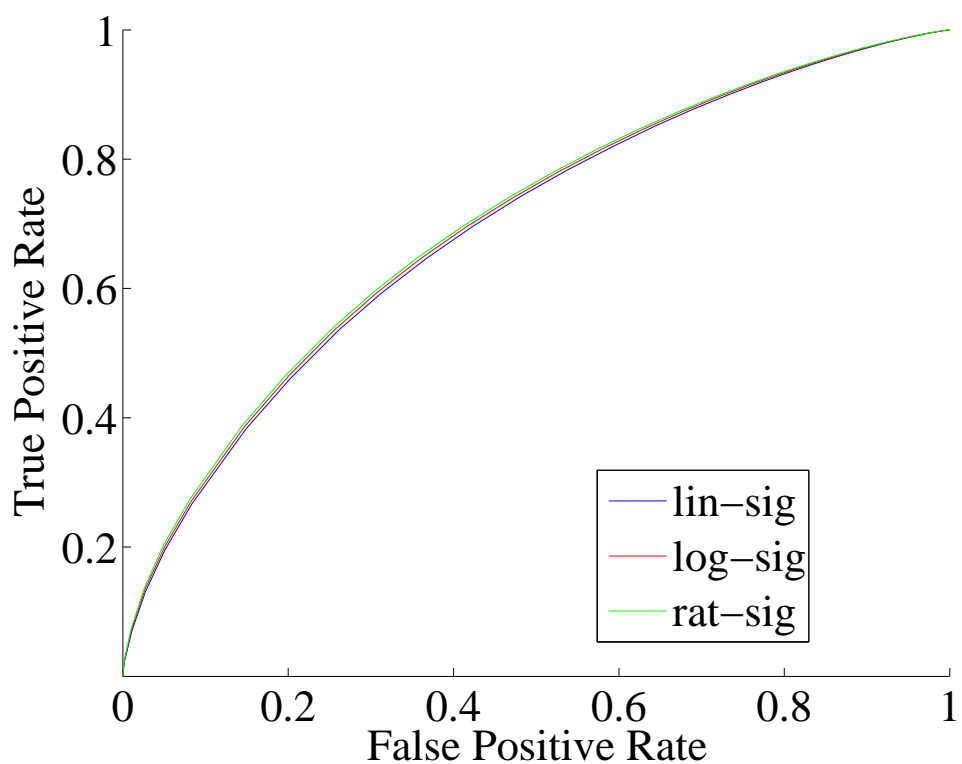


Рис. 43: ROC-кривые для функций оценки *linsig*, *logsig*, *ratsig*.

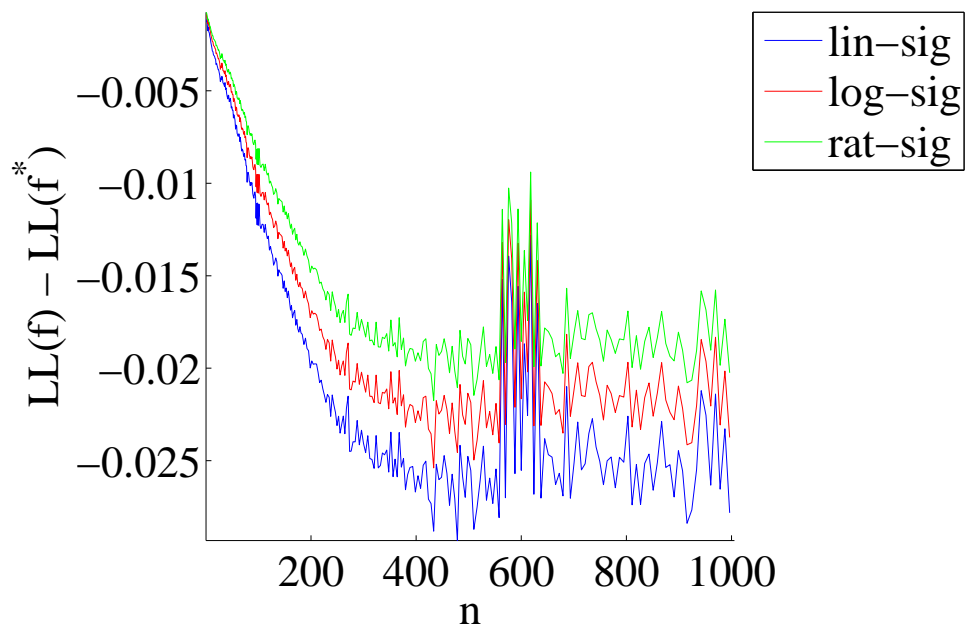


Рис. 44: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки *linsig*, *logsig*, *ratsig*.

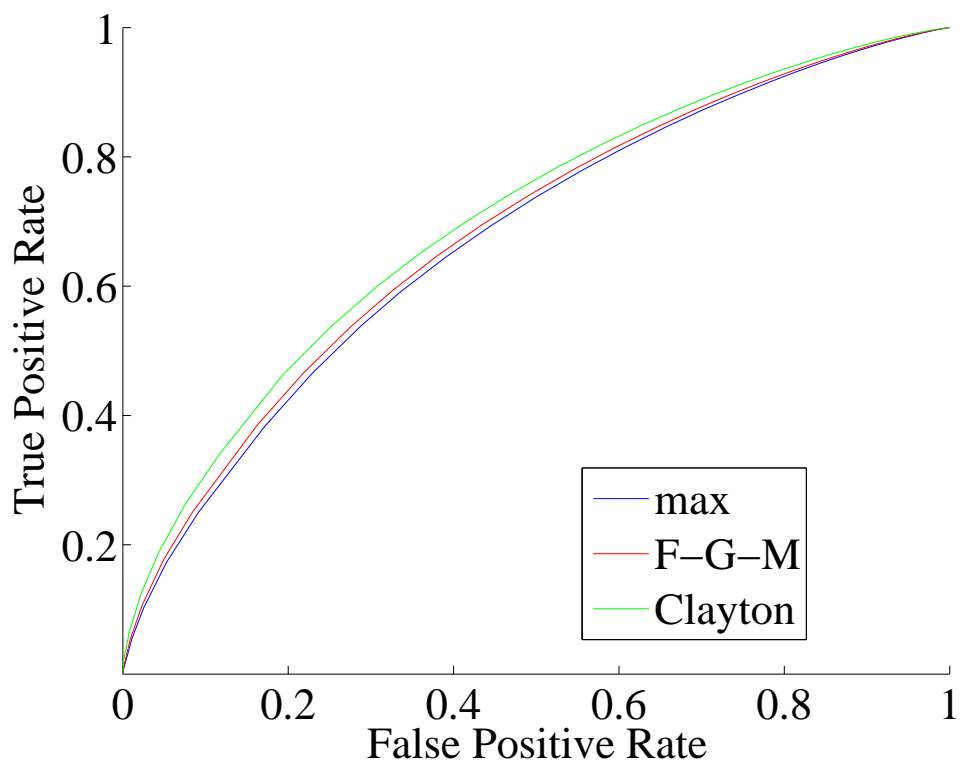


Рис. 45: ROC-кривые для функций оценки max , $F - G - M$, $Clayton$.

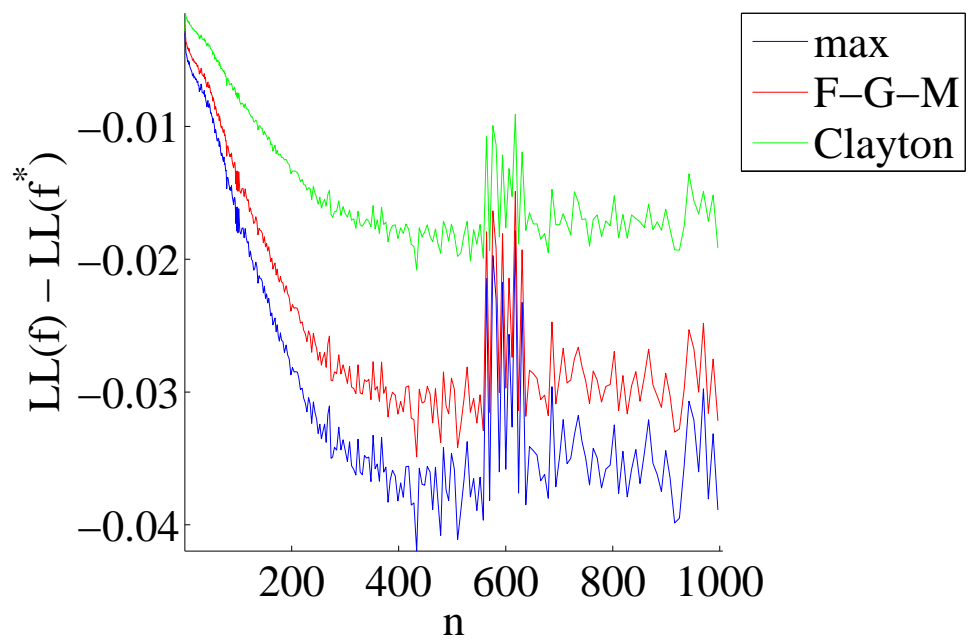


Рис. 46: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки max , $F - G - M$, $Clayton$.

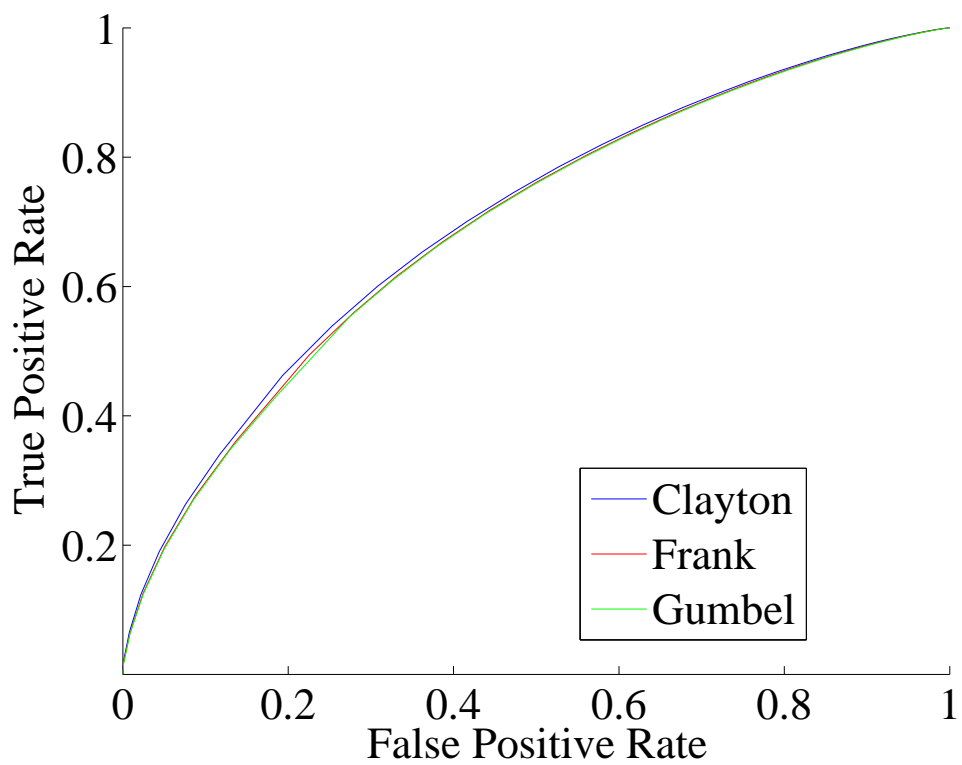


Рис. 47: ROC-кривые для функций оценки *Clayton*, *Frank*, *Gumbel*.

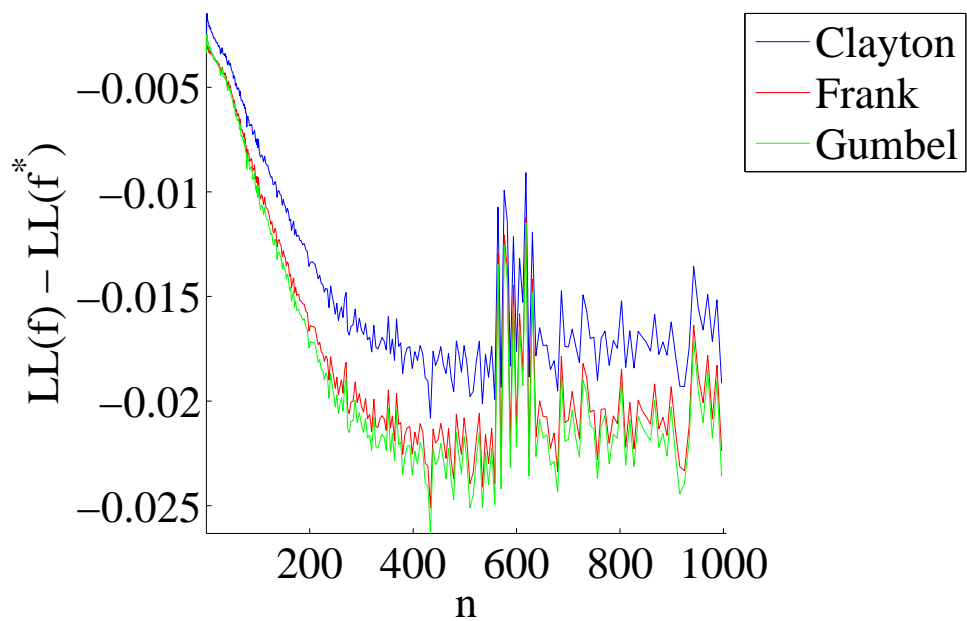


Рис. 48: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки *Clayton*, *Frank*, *Gumbel*.

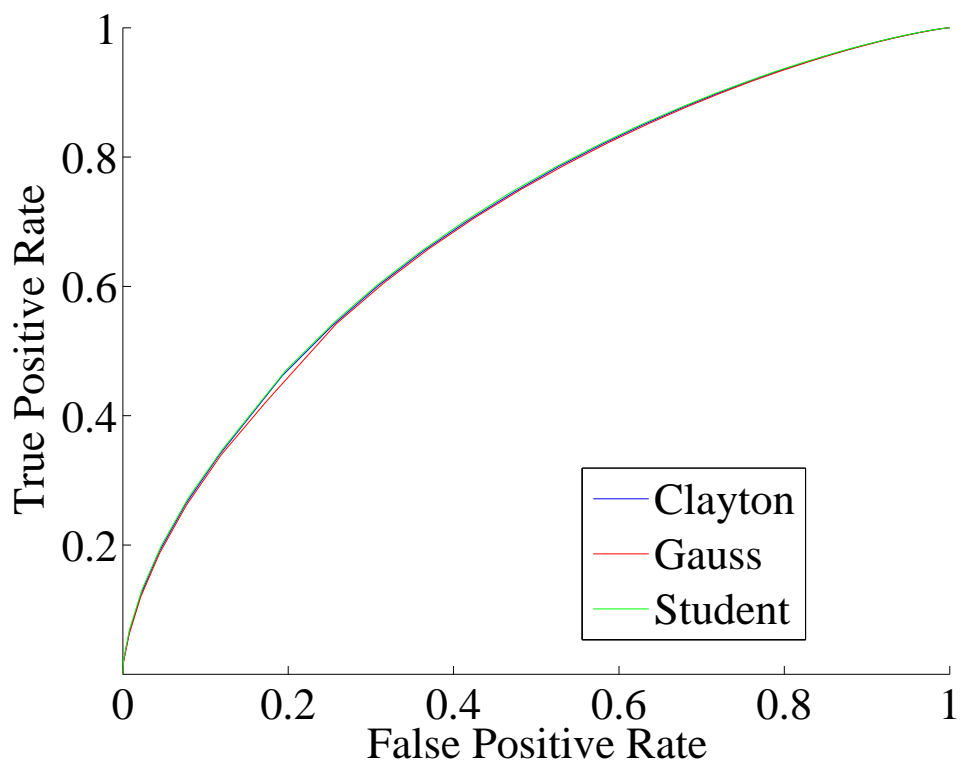


Рис. 49: ROC-кривые для функций оценки *Clayton*, *Gauss*, *Student*.

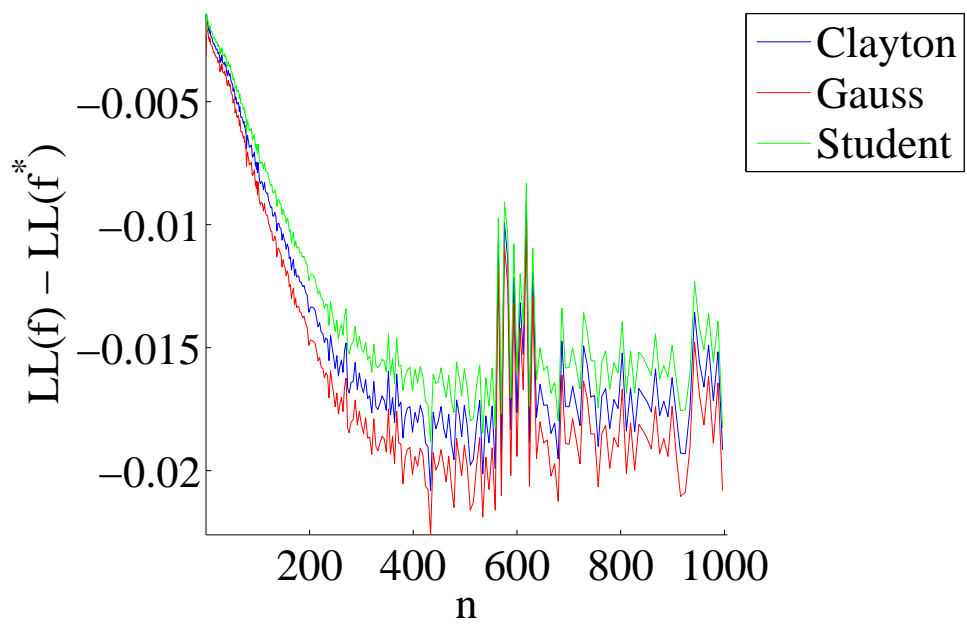


Рис. 50: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки *Clayton*, *Gauss*, *Student*.

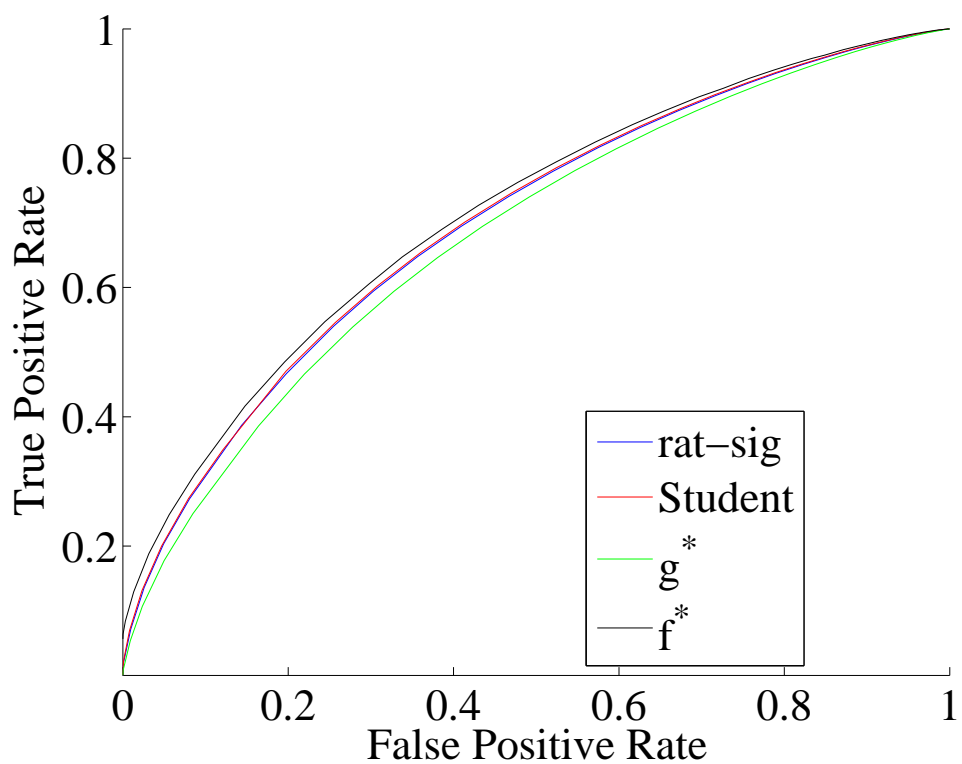


Рис. 51: ROC-кривые для функций оценки f^* , $ratsig$, $Student$, g^* .

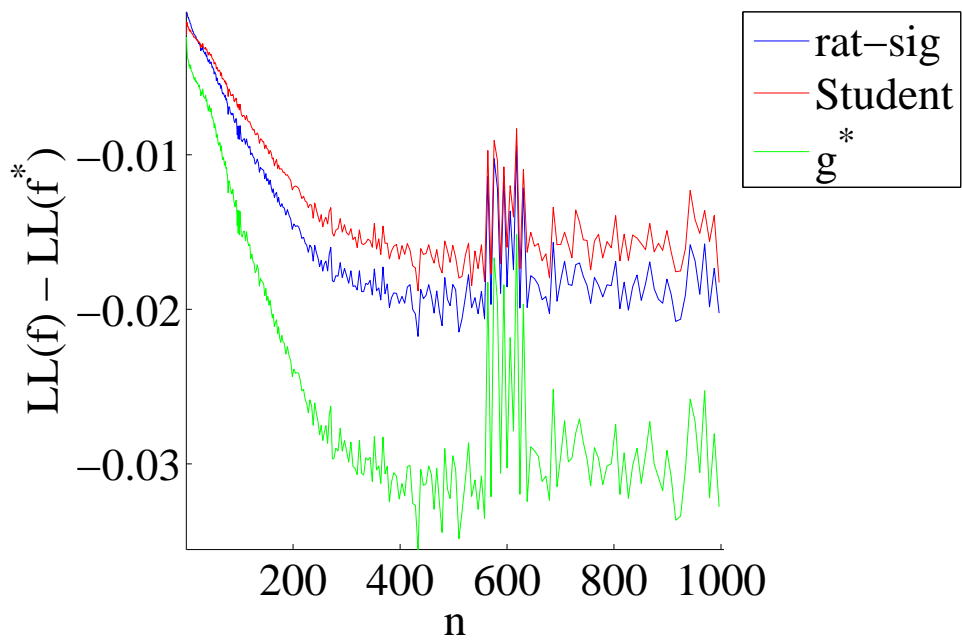


Рис. 52: Разности $\mathcal{L}(f) - \mathcal{L}(f^*)$ для функций оценки $ratsig$, $Student$, g^* .

6 Заключение

Рассмотрено два подхода к решению задачи. Функциональный подход позволяет выводить функцию оценки, соответствующие определённым содержательным требованиям, которые могут исходить, в том числе, из предметной области. Получаемая оценка аналитична и легко вычислима. Для оценки возможно выписать производные и проводить оптимизацию градиентными методами. Недостатком подхода является сильное предположение о функциональной зависимости. Стохастический подход ослабляет предположение о зависимости. Оценка однозначно определяется априорным совместным распределением пары вероятностей успеха и вычисляется Байесовскими методами. Недостатком подхода являются сложность вычисления оценки и сложность оптимизации параметров модели. Проведён вычислительный эксперимент на реальных данных, предоставленных компанией Яндекс. Получены верхние оценки значений функционала правдоподобия в рассматриваемом классе оценок и в редуцированном классе оценок. В рамках каждого подхода найдена функция оценки, превосходящая по качеству верхнюю оценку в редуцированном классе. Оценки стохастического подхода показали лучшее значение качества, чем оценки функционального подхода.

Список литературы

- [1] Ashish Agarwal, Kartik Hosanagar, and Michael D. Smith. Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets. *Journal of Marketing Research (JMR)*, 48(6):1057–1073, 2011.
- [2] Xingxing Wang, Shijie Lin, Dongying Kong, Liheng Xu, Qiang Yan, Siwei Lai, Liang Wu, Alvin Chin, Guibo Zhu, Heng Gao, Yang Wu, Danny Bickson, Yuanfeng Du, Neng Gong, Chengchun Shu, Shuang Wang, Kang Liu, Shuren Li, Jun Zhao, Fei Tan, and Yuanchun Zhou. Click-Through Prediction for Sponsored Search Advertising with Hybrid Models. 2012.
- [3] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the international conference on Web search and web data mining, WSDM '08*, pages 87–94, New York, NY, USA, 2008. ACM.
- [4] Matthew Richardson, Ewa Dominowska, and Robert Ragno. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web, WWW '07*, pages 521–530, New York, NY, USA, 2007. ACM.
- [5] Chia-Hua Ho An-Chun Liang Chun-Heng Huang Wei-Yuan Shen Jyun-Yu Jiang Ming-Hao Yang Ting-Wei Lin Ching-Pei Lee Perng-Hwa Kung Chin-En Wang Ting-Wei Ku Chun-Yen Ho Yi-Shu Tai I-Kuei Chen Wei-Lun Huang Che-Ping Chou Tse-Ju Lin Han-Jay Yang Yen-Kai Wang Cheng-Te Li Shou-De Lin Hsuan-Tien Lin Kuan-Wei Wu, Chun-Sung Ferng. A Two-Stage Ensemble of Diverse Models for Advertisement Ranking in KDD Cup 2012. 2012.
- [6] Juan Feng, Hemant K. Bhargava, and David M. Pennock. Implementing Sponsored Search in Web Search Engines: Computational Evaluation of Alternative Mechanisms. *INFORMS JOURNAL ON COMPUTING*, 19(1):137–148, January 2007.
- [7] Yuchen Zhang, Weizhu Chen, Dong Wang, and Qiang Yang. User-click modeling for understanding and predicting search-behavior. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '11*, pages 1388–1396, New York, NY, USA, 2011. ACM.

- [8] Ye Chen and Tak W. Yan. Position-normalized click prediction in search advertising. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '12, pages 795–803, New York, NY, USA, 2012. ACM.
- [9] Christian Genest and Anne C. Favre. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering*, 12(4):347–368, 2007.
- [10] Alexander J. McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: concepts, techniques and tools*. 2005.
- [11] Harry Joe. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419, June 2005.
- [12] James J. Xu Harry Joe. The Estimation Method of Inference Functions for Margins for Multivariate Models. *Technical Report no. 166, Department of Statistics, University of British Columbia*, 1996.