

# Модификация критерия Колмогорова–Смирнова для проекций степенного гамма-распределения с приложением в когнитивной биологии

Матушкина Ольга Дмитриевна, гр. 522

Санкт-Петербургский государственный университет  
Математико-механический факультет  
Кафедра статистического моделирования

Научный руководитель: к.ф.-м.н., д. Алексеева Н.П.

Рецензент: к.ф.-м.н., д. Коробейников А.И.



Санкт-Петербург

2013г.

# Возникновение задачи, выбор модели

- Экспериментальные данные: Forced swim test.
- Признаки: длительность (Duration) и время ожидания (Expectation) поведенческого акта.
- Сравнение индивидов (подопытных крыс) по оценкам параметров закона распределения Duration, Expectation.

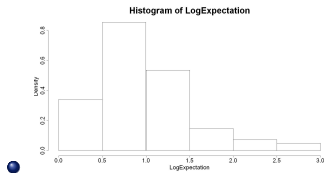


Рис. : Гистограмма LogExpectation акта floating для крысы 73

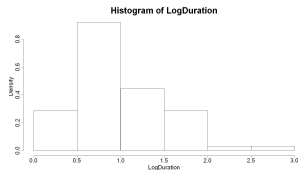


Рис. : Гистограмма LogDuration акта floating для крысы 58

Модель степенного гамма распределения и выбор проекции с более предпочтительными стат. свойствами.

- Проблемы критериев согласия при оцениваемых параметрах.
- Пусть  $x$  — сл. в. с плотностью распределения

$$g(x; \alpha, \beta) = \frac{1}{\beta} g\left[\frac{x - \alpha}{\beta}\right], \quad (1)$$

то есть  $\alpha$  — параметр сдвига,  $\beta$  — параметр масштаба.

## Теорема (С.Е. Antle, L.J. Bain, 1969)

*Если плотность имеет вид (1), то оценки макс. правдоподобия  $\hat{\alpha}$  и  $\hat{\beta}$  обладают свойством:  $\hat{\beta}_s = \hat{\beta} / \beta_0$ ,  $\hat{\alpha}_s = (\hat{\alpha} - \alpha_0) / \beta_0$  и  $\hat{\alpha}_p = (\hat{\alpha} - \alpha_0) / \hat{\beta}$  распределены независимо от  $\alpha_0$  и  $\beta_0$ .*

## Утверждение (J. Durbin, 1976)

### *Статистика критерия*

$$D(y) \equiv \sup_y \left| F(y; \hat{\alpha}, \hat{\beta}) - S_n(y) \right|,$$

*где  $S_n(y)$ —это эмпирическая ф. распределения, если плотность сл. в.  $Y$  зависит только от параметров сдвига и масштаба, распределена независимо от параметров распределения  $F(y; \alpha_0, \beta_0)$ .*

### **Требование к модели:**

- плотность зависит только от параметров сдвига и масштаба,
- оценки максимального правдоподобия (о.м.п.),
- критерий Колмогорова–Смирнова.

- Плотность распределения Вейбулла:

$$W(\gamma, \theta) = \frac{\gamma}{\theta} \left( \frac{x}{\theta} \right)^{\gamma-1} \exp\left(-\left(\frac{x}{\theta}\right)^{\gamma}\right), x \geq 0. \quad (2)$$

- Параметры формы  $\gamma$  и масштаба  $\theta$ .
- Пусть  $X \in W(\gamma, \theta)$ , тогда плотность  $Y = \log X$ :

$$\frac{1}{1/\gamma} e^{\frac{x - \log \theta}{1/\gamma}} \exp\left(e^{\frac{x - \log \theta}{1/\gamma}}\right). \quad (3)$$

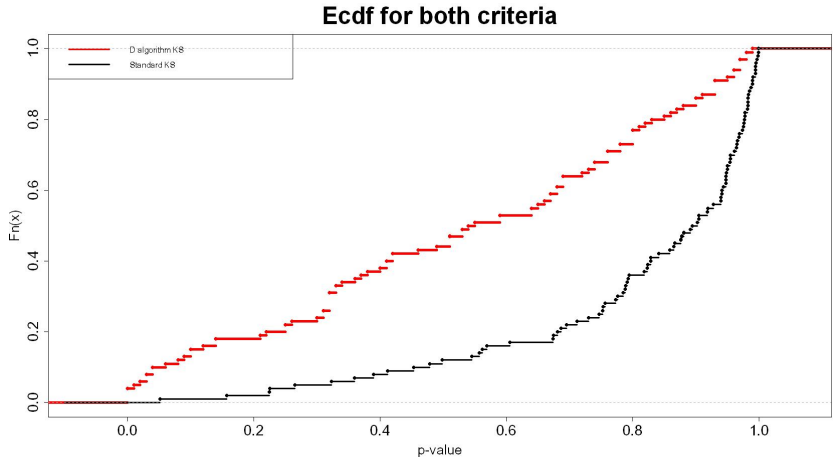
- Плотность гамма распределения:

$$\Gamma(k, \theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)}, \quad (4)$$

где  $x > 0$  и  $\Gamma(k)$ —это гамма-функция Эйлера.  
Не существует нужного преобразования.

# Модификация кр. Колмогорова–Смирнова для распределения Вейбулла

- Статистика критерия
  - ❶  $X = \{x_1, \dots, x_n\}$  — выборка.
  - ❷  $\hat{\theta}$  и  $\hat{\gamma}$  — о.м.п. для Вейбулла.
  - ❸  $Y = \log X$ .
  - ❹ Тестовая статистика:  $D(Y) \equiv \sup_y \left| F(y; \hat{\gamma}, \hat{\theta}) - S_n(y) \right|$ .
- p-value методом Монте–Карло.
- Распределение p-value при проверке нулевой гипотезы?
- Мощность критерия?



**Рис. :** Эмпирические функции распределения p-value при объеме выборки 100

- Объем выборки  $n$  от 100 до 1000 с шагом 100.
- Уровень значимости  $\alpha = 0.05$ .
- Критическое значение статистики  $D$  для  $n$  и  $\alpha$ .
- Сравнение полученного на модельной выборке значения  $D$  с критическим  $N$  раз.



# Нормальная $N(\mu, \sigma)$ и равномерная $U(\min, \max)$ альтернативы

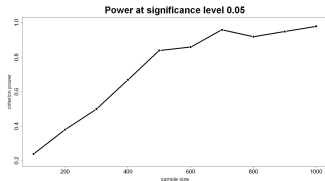


Рис. : Мощность критерия  $H_1 = N(\mu = 3, \sigma = 0.5)$

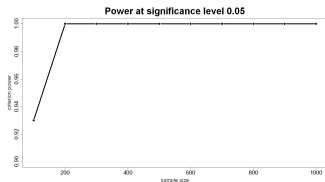


Рис. : Мощность критерия  $H_1 = U(\min = 0, \max = 1)$

- $Exp(\lambda) = \lambda e^{-\lambda x}, x \geq 0.$
- $\Gamma(1, 1/\lambda) = Exp(\lambda), Exp(\lambda) = W(1, 1/\lambda).$
- $k \neq 1$

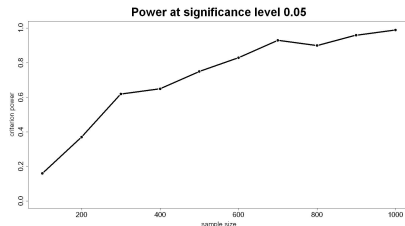


Рис. : Мощность критерия  $H_1 = \Gamma(k = 4, \theta = 1)$

- Forced swim test, поведенческий тест для оценивания уровня депрессии.
- Три генетические линии: Wistar (W), Wistar-Kyoto (WKY) и Spontaneously hypertensive rat (SHR).
- Переменные:
  - ① логарифм длительности акта, LogDuration,
  - ② логарифм время ожидания акта, LogExpectation.
- Анализируемые поведенческие акты:
  - ① floating (пассивное поведение)
  - ② swimming= swimming + paddling (активное поведение)
  - ③ grooming= wiping + headshakes (ориентированное на себя)

- Независимость как временных рядов, Mann-Kendall trend test для:
  - 1 LogDuration,
  - 2 LogExpectation.
- Переменные:
  - 1 Frequency (частота поведенческого акта),
  - 2 Mean(LogDuration) (среднее по логарифму от продолжительности акта),
  - 3 Mean(LogExpectation) (среднее по логарифму от времени ожидания акта).

Критерии:

- 1 однофакторный дисперсионный анализ,
  - 2 критерий Тьюки,
  - 3 критерий Краскела–Уоллиса.
- Различий между WKY и SHR не обнаружено. Тем самым показана необходимость привлечения статистической модели распределения Вейбулла.

- Согласие с моделью Вейбулла по поведенческим актам:
  - 1 floating и переменной LogDuration: у 71.5% животных,
  - 2 floating и переменной LogExpectation: у 71.5%,
  - 3 swimming и переменной LogDuration: у 51%,
  - 4 swimming и переменной LogExpectation: у 60%,
  - 5 grooming и переменной LogDuration: у 34%,
  - 6 grooming и переменной LogExpectation у 74.5%.

LogExpectation лучше описывается моделью, чем LogDuration.

- Новые переменные:
  - 1 shape,
  - 2 scale.
- Критерии:
  - 1 однофакторный дисперсионный анализ,
  - 2 критерий Тьюки,
  - 3 критерий Краскела–Уоллиса.
- Значимое различие между WKY и SHR:
  - 1 параметр scale по поведенческому акту swimming и переменной LogDuration,
  - 2 параметр scale по поведенческому акту grooming и переменной LogExpectation,
  - 3 параметр shape по поведенческому акту grooming и переменной LogExpectation.

- Реализация модификация критерия Колмогорова–Смирнова для распределения Вейбулла.
- преимущество Вейбулла перед другой проекцией — гамма.
- Проверка равномерности  $p$ -value при проверке нулевой гипотезы.
- Оценка мощности критерия при различных альтернативах.
- Возможность различить близкие распределение: Вейбулла и гамма.
- Новый подход: время ожидания и длительность акта.
- Значимое различие между WKY и SHR в модели распределения Вейбулла по отдельным актам.