1 Введение.

Замечание. Представим, что у нас есть какое-то множество с общим свойством. Все студенты ИТМО, все процессоры Intel, все автомобили автоваза. Мы хотим исследовать эти объекты, а их много и трудно перечислить. И если мы хотим исследовать то, что мы знаем, как получить, то проблем у нас нет (например, мы знаем, где найти каждого студента ИТМО, можно у каждого из них что-нибудь спросить). А вот изучить каждый процессор Intel — существенно сложнее. Короче, мы рассматриваем задачи, где объекты нельзя изучить каждый по отдельности.
Так, то есть у нас есть объекты. Множество всех таких объектов называется **генеральной совокупностью**. Мы не можем взять её всю, но можем выбрать какое-то подмножество (**выборку**). Хочется, зная параметры выборки, что-то сказать про генеральную совокупность. Если мы можем действительно сделать вывод, выборка называется **репрезентативной**. Что делать выборку репрезентативной?

- Размер. С одной стороны, если вам ответит половина человек, это при прочих равных лучше, чем если один процент.
- Дофига других факторов. Пример: если мы разместим опрос в интернете, то сколько бы людей не прошло опрос, выборка будет состоять из тех людей, которые пользуются интернетом в принципе и не только для работы.

Ладно, о'кей, как исследовать?

1. Ну, нужно собрать данные из какой-то выборки.
2. Нужно предобработать данные, потому что обычно они в сыром виде не готовы к изучению.
3. Выбор модели.
4. Прогонка данных через модель.
5. Интерпретация результата.

Пункты 1, 2 и 5 сильно зависят от предметной области, а вот 3 и 4 — то, что мы будет изучать.

Пример. Вот вопрос: бросается монетка много раз. Вопрос: честная ли монетка или нет? Решение примерно такое: если решка выпадает примерно половину раз, то монетка, вероятно, честная. Но вот насколько это **примерно** мы хотим считать успехом?

Пример. Есть люди, части которых дают лекарство, а другой части — плацебо. Вопрос: эффективно ли лекарство?

2 Простейшая модель выборки и друзья

Замечание. А если точнее «Простейшая модель выборки, эмпирическая функция распределения и её свойства, способы визуализации.»

Определение 1. Пусть имеются независимые и одинаково распределённые случайные величины $\xi_1, \ldots, \xi_n$ с функцией распределения F. Эта конструкция называется **простейшей моделью выборки**. n называется **объёмом выборки**. F называют **теоретической функцией распределения**.

Замечание. У нас известен на практике набор чисел (**реализация выборки** $x_1, \ldots, x_n$), а F мы не знаем.
А цель у нас — оценить F из реализации выборки.

Определение 2. **Эмпирической функцией распределения** называется

$$F_n(t) = \frac{\mu_n(t)}{n} \qquad \mu_n(t) = \sum_{i=1}^{n} \mathbb{I}(\xi_i \leqslant t)$$

Замечание. То есть $\mu(t)$ оценивает количество величин, не превосходящих t

Замечание. Эмпирическая функция распределения описывает дискретное распределение величины θ, в котором $P(\theta = x_i)$ равно количеству таких ξ_j, которые оказались равны x_i.

Свойство 2.1. *Заметим, что* $\mathbb{I}(\xi_1 \leqslant t)$ *имеет распределение* $\mathrm{Bern}(F(t))$, *а* $\mu_n(t)$ — $\mathrm{Bin}(n; F(t))$. *А значит*

$$\mathsf{E}\, F_n(t) = F(t) \qquad \mathsf{D}\, F_n(t) = \frac{1}{n}F(t)(1-F(t)) \qquad F_n(t) \xrightarrow{P} F(t)$$

Последнее — следствие ЗБЧ.

Замечание. $\mathsf{E}\, F_n(t) = F(t)$ называется **несмещённость**.

Замечание. Напоминание: сходимость по вероятности значит следующее:

$$P(|F_n(t) - F(t)| < \varepsilon) \to 1$$

Замечание. $F_n(t) \xrightarrow{P} F(t)$ называется **состоятельностью**. Ещё есть **сильная состоятельность** — сходимость почти наверное.

Свойство 2.2 (Асимптотическая нормальность)**.**

$$\frac{\mu_n(t) - nF(t)}{\sqrt{nF(t)(1-F(t))}} \xrightarrow{d} \mathrm{N}(0,1)$$

Это следствие ЦПТ.

Замечание. Также асимптотическую нормальность пишут как

$$\frac{\mu_n(t) - nF(t)}{\sqrt{nF(t)(1-F(t))}} \approx \mathrm{N}(0,1)$$

Рассуждение.

$$\begin{aligned} P(|F_n(t) - F(t)| < \varepsilon) &= P\left(\sqrt{n}\frac{|F_n(t) - F(t)|}{\sqrt{F(t)(1-F(t))}} < \frac{\sqrt{n}\varepsilon}{\sqrt{F(t)(1-F(t))}}\right) \approx \\ &\approx \Phi\left(\frac{\sqrt{n}\varepsilon}{\sqrt{F(t)(1-F(t))}}\right) - \Phi\left(-\frac{\sqrt{n}\varepsilon}{\sqrt{F(t)(1-F(t))}}\right) = \\ &= 2\Phi\left(\frac{\sqrt{n}\varepsilon}{\sqrt{F(t)(1-F(t))}}\right) - 1 \end{aligned}$$

Заметим следующий факт: $\sqrt{F(t)(1-F(t))} \leqslant \frac{1}{4}$. А значит выражение выше больше либо равно такого:

$$2\Phi(2\varepsilon\sqrt{n}) - 1 = \gamma \in (0;1)$$

Как найти ε?

$$\Phi(2\varepsilon\sqrt{n}) = \frac{\gamma+1}{2}$$

Ну так мы знаем такую штуку как квантиль!

$$2\varepsilon\sqrt{n} = q_{\frac{\gamma+1}{2}} \Leftrightarrow \varepsilon = \frac{q_{\frac{\gamma+1}{2}}}{2\sqrt{n}}$$

Отсюда что?

$$P\left(F(t) - \frac{q_{\frac{\gamma+1}{2}}}{2\sqrt{n}} < F_n(t) < F(t) + \frac{q_{\frac{\gamma+1}{2}}}{2\sqrt{n}}\right) \geqslant \gamma$$

Определение 3. В рассуждении выше $(F(t)-\varepsilon; F(t)+\varepsilon)$ — **асимптотический доверительный интервал** уровня γ для $F_n(t)$.

Рассуждение. Окей, пусть $t_0 < t_1 < t_2 < \dots < t_m$. Тогда

$$0 = F(t_0) \leqslant F(t_1) \leqslant \dots \leqslant F(t_m) = 1$$

Пусть $p_i = \Delta F(t_i) = F(t_i) - F(t_{i-1})$. И пусть $\mu_i = \Delta\mu_n(t_i) = \mu_n(t_i) - \mu_n(t_{i-1}) = \sum\limits_{j=1}^{n} \mathbb{I}(t_{i-1} < \xi_j \leqslant t_i)$. Тогда, несложно заметить,

$$\begin{pmatrix} \nu_1 \\ \vdots \\ \nu_m \end{pmatrix} \sim \mathrm{Poly}(n, p) \qquad p = \begin{pmatrix} p_1 \\ \vdots \\ p_m \end{pmatrix}$$

Вспомним, что $F_n(t_i) = \frac{\mu_n(t_i)}{n} = \frac{\nu_1 + \dots + \nu_i}{n}$. А значит $\Delta F_n(t_i) = \frac{\nu_i}{n}$. Пусть

$$W_n(t) = \sqrt{n}(F_n(t) - F(t)) \qquad \Delta W_n = \sqrt{n}(\Delta F_n - \Delta F) = \sqrt{n}\left(\frac{\nu_i}{n} - p_i\right) = \frac{\nu_i - np_i}{\sqrt{n}} \qquad \begin{pmatrix} \Delta W_1 \\ \cdots \\ \Delta W_n \end{pmatrix} = \nu^*$$

Тогда по многомерной ЦПТ $\nu^* \xrightarrow{d} N(\mathbb{0}; \Sigma)$, где Σ — матрица ковариаций ν.

Теорема 1 (Теорема Колмогорова)**.** *Пусть* $D_n = \sup\limits_{t \in \mathbb{R}} |F_n(t) - F(t)|$. *Тогда*

$$\lim_{n\to\infty} P(\sqrt{n} D_n \leqslant t) = K(t) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2t^2}$$

Теорема 2 (Теорема Гливенко — Кантелли)**.**

$$\sup_{t\mathbf{i}\mathbb{R}} |F_n(t - F(t)| \xrightarrow[n\to\infty]{a.s.} 0$$

Теорема 3 (Теорема Смирнова)**.** *Пусть* ξ_i *— одна выборка,* η_j *— другая выборка, и они независимы. Пусть*

$$D_{m,n} = \sum_{x \in \mathbb{R}} |F_n(x) - F_m(t)|$$

И пусть $F_\xi = F_\eta$. *Тогда*

$$\lim_{m,n\to\infty} P\left(\sqrt{\frac{mn}{m+n}} D_{m,n} \leqslant t\right) = K(t)$$

Визуализация .

Замечание. И как это визуализировать? Если выборка дискретна, есть такой вариант: можно посчитать, сколько раз каждое значение выпало в выборке и нарисовать частоты и значения. Если же выборка непрерывна, группировать можно по промежуткам.
Есть гистограмма: то же самое, только не график.
Но есть ещё способ...

Определение 4. Пусть мы взяли ξ_i и отсортировали:

$$\xi_{(1)} \leqslant \xi_{(2)} \leqslant \dots \leqslant \xi_{(n)}$$

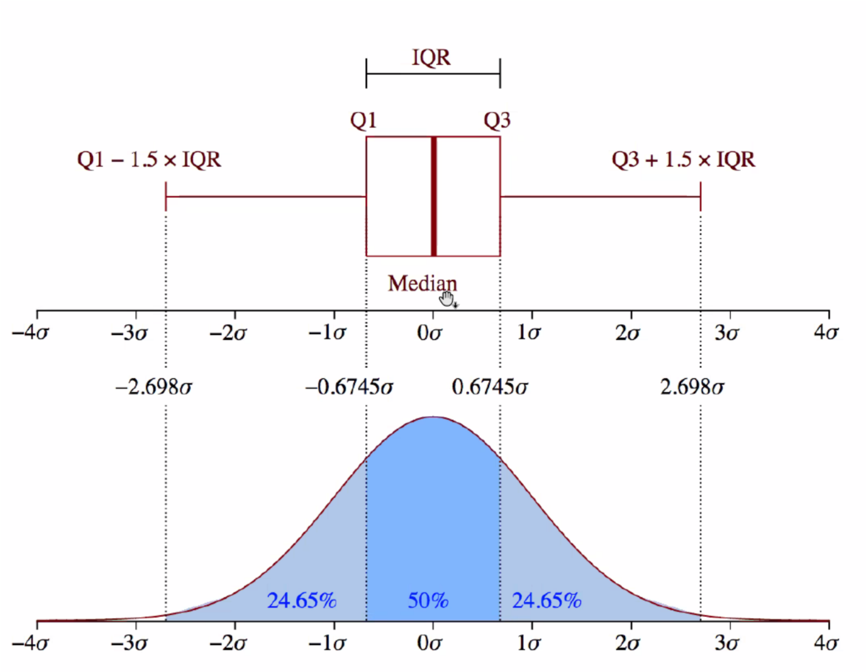
То есть $\xi_{(1)}$, например, это минимум всех ξ_i.
Так вот ряд $\xi_{(i)}$ называется **вариационным рядом**.

Определение 5. **Выборочная медиана** — это

$$\operatorname{med} \xi = \begin{cases} \frac{\xi_{(m)} + \xi_{(m+1)}}{2} & n = 2m \\ \xi_{(m+1)} & n = 2m + 1 \end{cases}$$

Также она называется **выборочной квантилью** порядка $\frac{1}{2}$.
Аналогично определяются выборочные квантили порядка $\frac{1}{4}$ и $\frac{3}{4}$. Они ещё соотвественно называются **верхним** и **нижнем квартилем**. Разница между ними — **межквартильный размах**.

Замечание. Так вот, оставшийся способ — box plot. Мы рисуем прямоугольник. Его середина по вертикали выделяется жирным и соотвествует медиане. Его вертикальные края соответствуют квартилям. И тут мы отмечаем значения выборки. При этом то, что по горизонтали находится дальше, чем полтора размаха, считается выбросами.



Выборочные моменты.

Определение 6. $\alpha_k = \mathsf{E}\,\xi_1^k$ — **k-тый теоретический момент**.

Определение 7. $\beta_k = \mathsf{E}(\xi_1 - \mathsf{E}\,\xi_1)^k$ — **k-тый центральный теоретический момент**.

Определение 8. Пусть у нас есть $g\colon \mathbb{R} \to \mathbb{R}$. Тогда

$$\overline{g(x)} = \frac{1}{n}\sum_{k=1}^{n} g(\xi_k)$$

Определение 9. $\widehat{\alpha_k} = \overline{\xi^k} = \frac{1}{n}\sum\limits_{j=0}^{n} \xi_j^k$ — **k-тый выборочный момент**.

Утверждение (несмещённость)**.**

$$\mathsf{E}\,\widehat{\alpha_k} = \alpha_k$$

Утверждение.

$$\mathsf{D}\,\widehat{\alpha_k} = \frac{1}{n}\mathsf{D}\,\xi_1^k = \frac{1}{n}(\mathsf{E}\,\xi_1^{2k} - (\mathsf{E}\,\xi^k)^2)$$

Утверждение (состоятельность)**.** *По ЗБЧ*

$$\widehat{\alpha_k} \xrightarrow{p} \alpha_k$$

Утверждение. *Из ЦПТ следует, что*

$$\sqrt{n}\frac{\widehat{\alpha_k} - \alpha_k}{\alpha_{2k} - \alpha_k^2} \approx \mathrm{N}(0; 1)$$

При этом

$$\sqrt{n}\frac{\widehat{\alpha_k}-\alpha_k}{\widehat{\alpha_{2k}}-\widehat{\alpha_k}^2}=\sqrt{n}\frac{\widehat{\alpha_k}-\alpha_k}{\alpha_{2k}-\alpha_k^2}\cdot\frac{\alpha_{2k}-\alpha_k^2}{\widehat{\alpha_{2k}}-\widehat{\alpha_k}^2}$$

Левая часть по распределению стремится у $\mathrm{N}(0;1)$*, а правая — по вероятности сходится к 1, так как*

$$\widehat{\alpha_{2k}}-\widehat{\alpha_k}^2\xrightarrow{p}\alpha_{2k}-\alpha_k^2$$

Определение 10. Также $\overline{\xi}$ называется **выборочным средним**.

Замечание. При выборочном среднем в знаменателе дроби выше написана дисперсия.

Определение 11. $\widehat{\beta_k}=\overline{(\xi-\overline{\xi})^k}=\frac{1}{n}\sum\limits_{j=1}^{n}(\xi_j-\overline{\xi})^k$ называется **k-тым центральным выборочным моментом**.

Определение 12. $\widehat{\beta_2}$ называется **выборочной дисперсией**, а $S_*=\sqrt{\widehat{\beta_2}}$ — **выборочным стандартным (среднеквадратичным) отклонением**.

Замечание. Выборочные моменты есть ни что иное как моменты, посчитанные относительно эмпирического распределения. Отсюда, напрмиер, $S_*^2=\overline{\xi^2}-\overline{\xi}^2$. И куча других формул вида

$$\mathsf{E}(\xi-\mathsf{E}\,\xi)^k=\mathrm{poly}(\mathsf{E}\,\xi,\ldots,\mathsf{E}\,\xi^k)$$

Утверждение. *Из последнего k-тый выборочный центральный момент является состоятельной оценкой k-того центрального теоретического момента.*

Отступление.

Утверждение. *Пусть Ξ_n — последовательность случайных векторов. Пусть $\sqrt{n}(\Xi_n-\mu)\xrightarrow{d}\mathrm{N}(\mathbb{0};\Sigma)$. Тогда $\Xi_n\xrightarrow{p}\mu$.*

Доказательство. Ну, $(\Xi_n-\mu)\frac{\sqrt{n}}{\sqrt{n}}$ по распределению сходится к нулю, а для вырожденных величин по распределению и по вероятности одно и то же. □

Замечание. Отсюда и далее градиент — строка.

Утверждение. *Пусть $\phi\in C^1(\mathbb{R}^m\to\mathbb{R})$. Тогда*

$$\sqrt{n}(\phi(\Xi_n)-\phi(\mu))\approx\sqrt{n}\nabla\phi(\mu)(\Xi_n-\mu)\to\mathrm{N}(0;\nabla\phi(\mu)\Sigma(\nabla\phi(\mu))^{\mathsf{T}})$$

Доказательство. Тейлор с остатком в Лагранже: $\phi(\Xi_n)=\phi(\mu)+\nabla\phi(\tilde{\mu})(\Xi_n-\mu)$. Если $n\to\infty$, то $\nabla\phi(\tilde{\mu})\to\nabla\phi(\mu)$. Отсюда

$$\phi(\Xi_n)-\phi(\mu)\approx\nabla\phi(\mu)(\Xi_n-\mu)$$

и

$$\mathsf{D}(\phi(\Xi_n)-\phi(\mu))\approx\mathsf{D}(\nabla\phi(\mu)(\Xi_n-\mu))=\mathsf{D}(\nabla\phi(\mu)\Xi_n)=\nabla\phi(\mu)\,\mathsf{D}\,\Xi_n(\nabla\phi(\mu))^{\mathsf{T}}$$

А отсюда

$$\sqrt{n}(\phi(\Xi_n)-\phi(\mu))\approx\sqrt{n}\nabla\phi(\mu)(\Xi_n-\mu)\to\mathrm{N}(0;\nabla\phi(\mu)\Sigma(\nabla\phi(\mu))^{\mathsf{T}})$$

□

Определение 13. Всё вышеописанное — дельта-метод.

Теорема 4. *Многомерная ЦПТ: $\sqrt{n}(\Xi_n-\alpha)\overline{d}\to\mathrm{N}(0;\Sigma)$, где $\alpha=(\alpha_1;\ldots;\alpha_k)$, $\Sigma=\mathsf{D}(\xi_1;\ldots;\xi_1^k)$.*

Теорема 5. *Пусть* $\Xi_n = \left(\overline{\xi}, \ldots, \overline{\xi^k}\right)$. *Пусть* $\phi \in C(\mathbb{R}^k \to \mathbb{R})$. *Пусть*

$$\sigma = \nabla\phi(\alpha)\Sigma(\nabla\phi(\alpha))^{\mathsf{T}}$$

Тогда

$$\sqrt{n}\frac{\phi(\Xi_n) - \phi(\alpha)}{\sigma} \xrightarrow{d} \mathrm{N}(0;1)$$

Кроме того если $\sigma(\alpha) \in C(\mathbb{R}^k)$, *то*

$$\sqrt{n}\frac{\phi(\Xi_n) - \phi(\alpha)}{\sigma(\Xi_n)} \xrightarrow{d} \mathrm{N}(0;1)$$

Утверждение (Упражнение)**.**

$$\sqrt{n}\frac{S_*^2 - \sigma^2}{\sqrt{\widehat{\beta_4} - S_*^4}} \approx \mathrm{N}(0;1)$$

Утверждение (Упражнение)**.**

$$\mathsf{E}\, S_*^2 = \frac{n-1}{n}\sigma^2$$

Определение 14.

$$S^2 = \frac{n}{n-1}S_*^2 = \frac{1}{n-1}\sum_{j=1}^{n}(\xi_i - \overline{\xi})^2$$

называется **исправленной (несмещённой) дисперсией**.

Определение 15. **Коэффициент ассиметрии** — это $\dfrac{\mathsf{E}(\xi - \mathsf{E}\,\xi)^3}{\sigma^3}$.

Определение 16. **Выборочный коэффициент ассиметрии** — это $\widehat{\gamma} = \dfrac{\widehat{\beta_3}}{S_*^3}$.

Определение 17. **Коэффициент эксцесса** — это $\dfrac{\mathsf{E}(\xi - \mathsf{E}\,\xi)^4}{\sigma^4} - 3$.

Определение 18. **Выборочный коэффициент эксцесса** — это $\dfrac{\widehat{\beta_4}}{S_*^4} - 3$.

Определение 19. Пусть у нас есть выборка $\xi_1, \ldots, \xi_n$ и другая выборка $\eta_1, \ldots, \eta_n$. **Выборочной ковариацией** называется

$$S_{*\xi\eta} = \frac{1}{n}\sum_j \xi_j\eta_j - \overline{\xi}\overline{\eta} = \frac{1}{n}\sum_j (\xi_j - \overline{\xi})(\eta_j - \overline{\eta})$$

Определение 20. Выборочным коэффициентом корреляции называется $\rho_n = \dfrac{S_{*\xi\eta}}{S_{*\xi}S_{*\eta}}$

3 Порядковые статистики.

Определение 1. **Вариационный ряд** — отсортированная по неубыванию выборка. Обозначение: $\xi_{(i)}$.
$\xi_{(k)}$ называется k-той порядковой статистикой.

Замечание. В некоторых книжках вариационный ряд — не просто отсортированная, но и сгруппированная по частоте.

Определение 2. **Квантилью** порядка α называется число q_α такое что

$$P(\xi \geqslant q_\alpha) \geqslant 1-\alpha \qquad\qquad P(\xi \leqslant q_\alpha) \geqslant \alpha$$

Утверждение. *Если функция распределения строго возрастает, q_α определяется единственным образом: $q_\alpha = F^{-1}(\alpha)$.*

Утверждение. *В противном случае квантилью можно считать либо $\sup\{x \mid F(x) \leqslant \alpha\}$, либо $\inf\{x \mid F(x) \geqslant \alpha\}$.*

Определение 3. **Выборочной квантилью** порядка 0 называется минимум, **выборочной квантилью** порядка q называется максимум, выборочной квантилью порядка $\alpha \in (0;1)$ называется вот что. Заметим, что

$$\exists k \in [0..n-1].\ \frac{k}{n} \leqslant \alpha < \frac{k+1}{n}$$

Тогда $\xi_{(k+1)}$ — **выборочная квантиль** порядка α.

Определение 4. q_0 называется **нижним квартилем**, $q_{1/4}$ — **первый (нижний) квартиль**, $q_{1/2}$ — **второй квартиль**, $q_{3/4}$ — **третий (верхний) квартиль**, q_1 — **четвёртый квартиль**.

Определение 5. **Межквартильный размах** (IQR) — $q_{3/4} - q_{1/4}$.

Определение 6. **Выборочной медианой** называется $\xi_{(m+1)}$, если $n = 2m+1$ или $\frac{\xi_{(m)}+\xi_{(m+1)}}{2}$.

Утверждение. *Что значит, что $P(\xi_{(k)} \leqslant t)$? Что*

$$P(\mu_n(t) \geqslant k) = \sum_{j=k}^{n} \binom{n}{k} (F(t))^j (1-F(t))^{n-j} = B(F(t);k;n-k+1)$$

Определение 7. **Обобщённой β-функцией** называется

$$B(z;a;b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int\limits_0^z t^{a-1}(1-t)^{b-1}\mathrm{d}t$$

Утверждение. *Пусть $p = F'$. Тогда плотность k-той порядковой статистики равна*

$$\frac{\mathrm{d}}{\mathrm{d}t} B(F(t);k;n-k+1) = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} (F(t))^{k-1}(1-F(t))^{n-k} p(t)$$

Утверждение. *Совместная плотность двух порядковых статистик равна*

$$g(x_1;x_2) = \frac{n!}{(k-1)!(r-k-1)!(n-r)!} (F(t))^{k-1}(1-F(t))^{r-k-1} p(x_1)p(x_2)$$

Утверждение. *Совместная плотность всех порядковых статистик равна*

$$g(x_1;x_2;\ldots;x_n) = n!p(x_1)p(x_2)\cdots p(x_n)$$

Определение 8. Средним членом вариационного ряда называется член с номером $k(n)$: $\frac{k(n)}{n} \to \mathrm{const} \in (0;1)$.
Крайний же член вариационного ряда — член с номером r (где r ограничено) или с номером $n-1+s$, где s ограничено.

Теорема 1 (Об асимптотике среднего члена вариационного ряда)**.** *Пусть $\alpha \in (0;1)$, q_α — теоретическая квантиль порядка α, $p \in C^{(1)}(V_{q_\alpha})$ —теоретическая плотность. И пусть $p(q_\alpha) > 0$.*
Тогда $\sqrt{n}p(q_\alpha)\frac{\xi_{(\lfloor n\alpha \rfloor)} - q_\alpha}{\sqrt{\alpha(1-\alpha)}} \xrightarrow{d} \mathrm{N}(0;1)$

Доказательство. Идея доказательства: пусть $k = \lfloor n\alpha \rfloor$. Пусть $g(x) = \sqrt{n}p(q_\alpha)\frac{x-q_\alpha}{\sqrt{\alpha(1-\alpha)}}$. Напишем плотность этого:

$$p_{g(\xi_{(k)})}(t) = p_{\xi_{(k)}}(g^{-1}(t)) \cdot g^{-1}(t)'_t$$

Дальше надо расписать всё по Стирлингу, и в пределе будет что надо. □

Пример. Пусть у нас Cauchy$(\mu; 1)$, $\alpha = \frac{1}{2}$. Тогда

$$p(\mu)\sqrt{n}2(\xi_{(\lfloor \frac{n}{2} \rfloor)} - \mu) = \frac{1}{\pi}\frac{1}{1+(\mu-\mu)^2}2\sqrt{n}(\xi_{(\lfloor \frac{n}{2} \rfloor)} - \mu)$$

Теорема 2 (Об асимптотике крайних членов вариационного ряда)**.** *Пусть p — плотность. Тогда*

$$nF(\xi_{(r)}) \xrightarrow{d} \Gamma(r; 1)$$

$$nF(\xi_{(n+1-s)}) \xrightarrow{d} \Gamma(s; 1)$$

Кроме того $\Gamma(r; 1)$ и $\Gamma(s; 1)$ независимы

4 Постановка задачи точечного оценивания параметров.

Замечание. Пусть у нас $\xi_1; \ldots; \xi_n$ имеют распределение θ, где $\theta \in \Theta \subset \mathbb{R}^d$. И θ — фиксированный неизвестный вектор.
Наша цель — оценить θ в виде $\widehat{\theta} = \widehat{\theta}(\xi_1; \ldots; \xi_n)$ (в виде функции от выборки).

Определение 1. Статистика — измеримая функция от выборки.

Замечание. Ещё есть Байесовская постановка задачи: там θ — случайная величина из известного распределения Θ.

Определение 2. Говорят, что $\widehat{\theta}$ — **состоятельная оценка** θ, если $\widehat{\theta} \xrightarrow{d} \theta$.

Определение 3. Смещением называется $\operatorname{bias}(\widehat{\theta}) = \mathsf{E}\,\widehat{\theta} - \theta$.

Определение 4. Если смещение равно нулю, оценка называется **несмещённой**.

Определение 5. Если смещение стремится к нулю, оценка называется **асимптотически несмещённой**.

Определение 6. Оценка $\widehat{\theta}$ **асимптотически нормальна**, если

$$\sqrt{n}(\widehat{\theta} - \theta) \to \mathrm{N}(0; \Sigma)$$

Пример. Пусть у нас есть Bern(p), и мы не знаем p. Каким мы хотим его взять? Ну, например $\overline{\xi}$. Тогда по ЗБЧ она состоятельна, очевидно несмещена и по ЦПТ асимптотически нормальна.
Пример. Пусть у нас $\mathrm{N}(\mu; \sigma^2)$. И пусть μ мы знаем, а σ^2 — нет. Кто мы берём? Например, S_*^2. Но увы, эта оценка смещена. Только асимптотическая несмещённость есть.

Определение 7. Будем говорить, что $\widehat{\theta}_1$ оптимальнее (эффективнее) $\widehat{\theta}_2$, если

$$\mathrm{MSE}(\widehat{\theta}_1) < \mathrm{MSE}(\widehat{\theta}_2)$$

Где

$$\mathrm{MSE}(\widehat{\theta}) = \mathsf{E}\,\|\widehat{\theta} - \theta\|^2 = \mathsf{E}(\widehat{\theta} - \theta)^{\mathsf{T}}(\widehat{\theta} - \theta)$$

Замечание. В некоторых книжках эти понятия различны, но у нас они будут одним.

Утверждение. *Если $\widehat{\theta}$ несмещена*

$$\mathrm{MSE}(\widehat{\theta}) = \operatorname{tr}(\mathsf{D}\,\widehat{\theta}) + \|\operatorname{bias}\widehat{\theta}\|^2$$

Доказательство. Тут надо воспользоваться тем, что $\theta = \mathsf{E}\,\widehat{\theta}$. □