3 Доверительные беседы о доверительных интервалах и множествах

3.1 Доверительные интервалы

3.1.1 Метод центральной функции

Будем рассматривать одномерные параметры $\theta \in \mathbb{R}$.

Напомним, что доверительным интервалом называют пару статистик $\widehat{\theta}_1$, $\widehat{\theta}_2$, таких, что при всех θ выполнено равенство

$$\mathbf{P}_\theta(\theta \in (\widehat{\theta}_1, \widehat{\theta}_2)) = 1 - \alpha,$$

где $1 - \alpha$ — заданное число, называемое *уровнем доверия*. Можно требовать взамен выполнение неравенства $\geq$ или сходимости $\to$ при $n \to \infty$ (в последнем случае интервал называется асимптотическим).

Основным методом для их построения является метод центральной функции:

- выбираем некоторую статистику T, например, достаточную;
- находим некоторую функцию $g(T(X_1, \ldots, X_n); \theta)$, монотонную по переменной θ и имеющую распределение, независящее от параметра;
- находим отрезок [a,b], в котором $g(T(X_1, \ldots, X_n); \theta)$ лежит с вероятностью $1 - \alpha$;
- обращаем функцию g по второй переменной и находим $\widehat{\theta}_1 = g^{-1}(T(x_1, \ldots, x_n); \cdot)(a)$, $\widehat{\theta}_2 = g^{-1}(T(x_1, \ldots, x_n); \cdot)(b)$.

Полученные статистики и есть границы интервала. Для нахождения указанных a, b используют так называемые квантили:

Определение 1. α-квантилью распределения F называют $F^{-1}(\alpha)$.

Если $g(T(X_1, \ldots, X_n); \theta)$ имеет ф.р. F с квантилями x_α, то

$$\mathbf{P}\left(g(T(X_1, \ldots, X_n); \theta) \in (x_\beta, x_{1-\alpha-\beta}]\right) = 1 - \alpha.$$

Значит,

$$\left[g^{-1}_{T(X_1,\ldots,X_n)}(x_\beta), g^{-1}_{T(X_1,\ldots,X_n)}(x_{1-\alpha-\beta})\right]$$

будет требуемым доверительным интервалом, где обратная функция к g берется по второй переменной при фиксированной первой.

Вопрос 1. Как выглядит доверительный интервал на основе $\max X_i$, где $X_i \sim R[0, \theta]$?

Функция g особенно удобно находится в случае, если плотность f_T является плотностью сдвига-масштаба, то есть плотность $f_T((x - a(\theta))/\sigma(\theta))/\sigma(\theta)$ не зависит от θ при некоторых $a(\theta)$, $\sigma(\theta)$. Тогда $g(T; \theta)$ можно взять вида $(T - a(\theta))/\sigma(\theta)$.

Статистику T обычно выбирают достаточной (если есть одномерная достаточная статистика), чтобы не терять информацию о параметре.

Параметр β обычно выбирают таким образом, что длина интервала минимальна. Здесь помогает следующая теорема

Теорема 1. *Если $f(x)$ унимодальная плотность (то есть имеющая единственный локальный максимум x^*), $a < x^* < b$, $f(a) > 0$, $f(b) > 0$ и*

$$\int_a^b f(x)dx = 1 - \alpha, \tag{1}$$

то минимальная длина $b - a$ будет достигаться при $f(a) = f(b)$ среди интервалов, удовлетворяющие (1).

3.1.2 Метод обращения функции

В случае, если параметр θ глубоко вшит в структуру распределения (например, параметры биномиального или пуассоновского распределения, параметры формы гамма-распределения или параметры бета-распределения), предыдущий метод не работает, однако, нам помогает следующая теорема:

Теорема 2. *Пусть T имеет ф.р., монотонно возрастающую по θ, $\widehat{\theta}_1(x)$, $\widehat{\theta}_2(x)$ определяются так, что*

$$\mathbf{P}_{\widehat{\theta}_1}(T \le x) = \beta, \quad \mathbf{P}_{\widehat{\theta}_2}(T \ge x) = \alpha - \beta.$$

Тогда $(\widehat{\theta}_1(T), \widehat{\theta}_2(T))$ является доверительным интервалом с уровнем доверия не менее $1-\alpha$.

В случае убывающей по θ функции интервал будет иметь границы, идущие в обратном порядке.

Вопрос 2. Показать, что для $X_i \sim Poiss(\lambda)$

$$(\widehat{\theta}_1, \widehat{\theta}_2), \quad \widehat{\theta}_1 = \frac{1}{2n} y_{1-\alpha/2, 2\sum_{i=1}^n X_i}, \quad \widehat{\theta}_2 = \frac{1}{2n} y_{\alpha/2, 2\sum_{i=1}^n X_i + 2}$$

является доверительным интервалом для λ.

3.1.3 Асимптотические доверительные интервалы

Построение точных доверительных интервалов достаточно трудоемко, а вот асимптотические доверительные интервалы можно строить на основе любой асимптотически нормальной оценки.

Если оценка $\widehat{\theta}$ является асимптотически нормальной с непрерывной асимптотической дисперсией $\sigma(\theta)$:

$$\sqrt{n}\frac{\widehat{\theta}(X_1, \ldots, X_n) - \theta}{\sigma(\theta)} \xrightarrow{d} Z \sim \mathcal{N}(0,1), \ n \to \infty,$$

то мы можем построить асимптотический доверительный интервал

$$\left(\widehat{\theta}(X_1, \ldots, X_n) - \frac{z_{1-\alpha/2}\sigma(\widehat{\theta}(X_1, \ldots, X_n))}{\sqrt{n}}, \widehat{\theta}(X_1, \ldots, X_n) + \frac{z_{1-\alpha/2}\sigma(\widehat{\theta}(X_1, \ldots, X_n))}{\sqrt{n}}\right)$$

Вопрос 3. Почему это действительно асимптотический доверительный интервал уровня доверия $1-\alpha$?

Вопрос 4. Почему для нормального распределения имеет смысл брать симметричный интервал с квантилями уровня $\alpha/2$ и $1-\alpha/2$?

Так если $\widehat{\theta}$ — ОМП для θ в сильно регулярной модели, то ее асимптотическая дисперсия есть $1/I(\theta)$ — величина, обратная к информации Фишера, которую можно состоятельно оценить величиной $I(\widehat{\theta})$. Отсюда мы можем построить доверительный интервал вида

$$\left(\widehat{\theta} - \frac{z_{1-\alpha/2}}{\sqrt{n}\sqrt{I(\widehat{\theta})}}, \widehat{\theta} + \frac{z_{1-\alpha/2}}{\sqrt{n}\sqrt{I(\widehat{\theta})}}\right).$$

Может показаться, что для оценки методом моментов доверительный интервал строится сложнее. Но здесь нам на помощь приходит дельта-метод. Асимптотическую дисперсию оценки $\overline{X^k}$ как оценки $\mu_k(\theta)$ мы знаем — это $\mathbf{D}_\theta X_k = \mu_{2k}(\theta) - \mu_k^2(\theta)$. Следовательно, если ОММ для θ есть $f(\overline{X^k})$, то ее асимптотическая дисперсия $\sigma^2(\theta) = (\mu_{2k}(\theta) - \mu_k^2(\theta))(f'(\mu_k(\theta)))^2$, то есть доверительный интервал будет иметь вид

$$\left(f(\overline{X^k}) - \frac{z_\beta \sigma(f(\overline{X^k}))}{\sqrt{n}}, f(\overline{X^k}) - \frac{z_\gamma \sigma(f(\overline{X^k}))}{\sqrt{n}}\right).$$

3.2 Доверительные множества или эллипс против прямоугольника

3.2.1 Доверительный эллипсоид для нормальных оценок

Для векторного параметра $\theta = (\theta_1, \ldots, \theta_k)$ нам уже не очень помогают доверительные интервалы для каждой θ_i, поскольку знание того, что каждая из $\theta_1, \ldots, \theta_k$ лежит в своем интервалом с заданной вероятностью не дает нам возможность выписать вероятность того, что все параметры одновременно попадут в соответствующий параллелипипед.

Можно, конечно, выделить по каждой оси $1 - \alpha/k$ интервал, тогда вероятность того, что хоть одна из θ_i не попадет в свой интервал, будет не больше α. Но этот параллелипипед будет слишком большим и в ряде случаев может быть значительно улучшен. Поэтому рассматривают доверительное множество $A(X_1, \ldots., X_n) \subset \mathbb{R}^k$, которое накроет мой параметр с вероятностью $1 - \alpha$, то есть

$$\mathbf{P}(\vec{\theta} \in A(X_1, \ldots, X_n)) = 1 - \alpha.$$

Для многомерной нормальной выборки $(X_1, \ldots, X_k) \sim \mathcal{N}(\vec{\mu}, \Sigma)$, мы можем утверждать, что $(\vec{X} - \vec{\mu})^{\mathrm{T}}\Sigma^{-1}(\vec{X} - \vec{\mu})$ будет вектором с распределением χ^2_k, поскольку $A^{-1}(\vec{X} - \vec{\mu}) \sim \mathcal{N}(0, E)$, где $A^{\mathrm{T}}A = \Sigma$. Но тогда мы можем сказать, что выборка попадает в эллипсоид с центром $\vec{\mu}$

$$(\vec{X} - \vec{\mu})^{\mathrm{T}}\Sigma^{-1}(\vec{X} - \vec{\mu}) \leq x$$

с вероятностью $F_{\chi^2_k}(x)$.

Вопрос 5. Показать, что это действительно так.

3.2.2 Доверительный эллипсоид для ОМП

Таким образом, мы можем строить асимптотические доверительные множества на основе векторных асимптотически нормальных оценок. Для ОМП в роли асимптотической дисперсии $\Sigma(\theta)$ будет выступать $I^{-1}(\theta)$, где $I(\theta)$ — информационная матрица Фишера с элементами

$$cov\left(\frac{\partial}{\partial\theta_i}\ln f_\theta(X), \frac{\partial}{\partial\theta_j}\ln f_\theta(X)\right).$$

$$\sqrt{n}(\widehat{\theta} - \vec{\theta}) \xrightarrow{d} \mathcal{N}(0, \Sigma(\theta)).$$

Поэтому мы имеем доверительный эллипсоид для $\vec{\theta}$, заданный соотношением

$$(\widehat{\theta} - \vec{\theta})^{\mathrm{T}} I(\widehat{\theta})(\widehat{\theta} - \vec{\theta}) \leq y_{1-\alpha},$$

где y — квантиль χ^2_k.

Информационная матрица может быть записана как $-E_\theta G$, где G — матрица Гессе для логарифмической функции правдоподобия. Поэтому вычисление оценки $\Sigma(\widehat{\theta})$ может быть сделано с помощью подсчета гессиана при нахождении ОМП.

В R при поиске ОМП с помощью nlm можно указать параметр hessian = TRUE. У выданной функцией nlm величины out параметр out$hessian будет содержать искомую оценку для матрицы Фишера (с точностью до знака).

Аналогичным образом scipy.optimize.minimize содержит метод hess, который позволяет подсчитать гессиан.

3.2.3 Доверительный эллипсоид для ОММ

Аналогично строится доверительное множество на основе ОММ, только для подсчета матрицы ковариации придется использовать многомерный Дельта-метод и ЦПТ для векторов. Более конкретно, если $\Sigma_1(\theta)$ — матрица ковариации вектора $(X, X^2, \ldots, X^k)$, g — отображение, такое что $g(\mu_1(\theta), \ldots, \mu_k(\theta)) = \theta$, а $J(\theta)$ — его матрица Якоби в точке $\mu_1(\theta), \dots, \mu_k(\theta)$, то для $\widehat{\theta} = g(\overline{X}, \overline{X^2}, \dots, \overline{X^k})$ выполнено соотношение

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} Z \sim \mathcal{N}(0, J^{\mathrm{T}}(\widehat{\theta})\Sigma_1(\widehat{\theta})J(\widehat{\theta})).$$

Соответственно, доверительный эллипсоид имеет вид

$$(\widehat{\theta} - \theta)^{\mathrm{T}}\Sigma^{-1}(\widehat{\theta})(\widehat{\theta} - \theta) \leq y_{1-\alpha}/n,$$

где $\Sigma = J\Sigma_1 J^{\mathrm{T}}$.

Задача 1. Построить асимптотический доверительный эллипс на основе ОМП для $X_i \sim \Gamma(\theta_1, \theta_2)$ на выборке размера 50 при $\theta_1 = 2$, $\theta_2 = 2$. Обратную матрицу к A можно построить с помощью функции solve(A).

В R доверительные эллипсы нормального распределения удобно рисовать с помощью функции ellipse пакета mixtools. Так функция ellipse(mu,Sigma,alpha = .05, npoints = 50, newplot = TRUE, type = "l") строит доверительный эллипс уровня alpha для нормального $\mathcal{N}(mu, Sigma)$ распределения.

3.2.4 Доверительный эллипсоид и Дельта-метод

Мы можем использовать дельта-метод для любой асимптотически нормальной оценки, зависящей от оценок с уже известным совместным распределением: если $\widehat{\theta}$ асимптотически нормальна для $\vec{\theta}$ с асимптотической матрицей ковариации $\Sigma(\theta)$, то $h(\widehat{\theta})$, $h : \mathbb{R}^m \to \mathbb{R}^k$ будет асимптотически нормальной оценкой $h(\vec{\theta})$ с асимптотической дисперсией

$$\Sigma_J(\theta) = J(\theta)\Sigma(\theta)J^t(\theta), \quad J = \left(\frac{\partial h_i}{\partial \theta_j},\ i, j\right).$$

Если матрица якоби J и матрица Σ непрерывно зависят от θ, то можно подставить туда оценку $\widehat{\theta}$ и построить доверительный эллипсоид

$$(h(\widehat{\theta}) - h(\vec{\theta}))\Sigma_J^{-1}(\widehat{\theta})(h(\widehat{\theta}) - h(\vec{\theta}))^t \leq y_{1-\alpha},$$

где $y_{1-\alpha}$ — квантиль χ^2_k.

3.3 Бутстрэп или как вытаскивать себя из болота за тесемки сапогов

3.3.1 Основная идея

Построение доверительного интервала или множества фактически связано с вопросом оценки "типичного" разброса нашей статистики вокруг настоящего параметра. Было бы удобно знать настоящее распределение выборки, тогда мы могли бы сгенерировать множество выборок из него, для каждой из них посчитать статистику и отсюда узнать как часто статистика отклоняется на то или иное расстояние от нашего параметра.

К сожалению, распределения мы не знаем, зато можем попробовать оценить его.

Итак, две основных идеи таковы:

- Если мы знаем функцию распределения какой-то выборки, то с помощью метода Монте-Карло мы можем приближенно найти моменты функций от наших величин, генерируя выборки из нашего распределения.

- Если у нас есть "хорошая оценка" $\widehat{\theta}$ неизвестного параметра θ, а семейство распределений непрерывно, то $F_{\widehat{\theta}}$ близко к F_θ

Метод бутстрэпа основывается на этих двух идеях. Он называется в честь ремешков на обуви в связи с идиомой, означающей "вытянуть себя из трясина за ремешки на обуви" (мы в таком случае вспоминаем косичку барона Мюнхгаузена). Этот метод позволяет нам улучшить оценку с помощью самой этой оценки.

3.3.2 Бутстрэп для исправления смещения оценок

Сперва отклонимся от построения доверительных интервалов и посмотрим на смещенные оценки.

Пример 1. Итак, пусть оценка $\widehat{\theta}$ достаточно близка к θ, но обладает небольшим смещением

$$a(\theta) = \mathbf{E}_\theta \widehat{\theta}(X_1, \ldots, X_n) - \theta.$$

Пусть она на нашей реализации выборки приняла значение $\widehat{\theta}(x_1, \ldots, x_n)$. Тогда мы можем взять множество выборок $Y_{i,1}, \ldots, Y_{i,n}$, $i = 1, \ldots, m$ из $F_{\widehat{\theta}(x_1,\ldots,x_n)}$ и подсчитать по ним

$$\widetilde{\theta}(Y) = \frac{1}{m} \sum_{i=1}^{m} \widehat{\theta}(Y_{i,1}, \ldots, Y_{i,n}).$$

Тогда смещение $\widetilde{\theta}(Y) - \widehat{\theta}(X_1, \ldots, X_n)$ при больших m близко к $a(\widehat{\theta})$. В свою очередь мы можем ожидать, что $a(\widehat{\theta})$ близко к $a(\theta)$. Таким образом, мы можем взять оценку

$$\widehat{\theta}(x_1, \ldots, x_n) - a(\widehat{\theta}) = 2\widehat{\theta}(x_1, \ldots, x_n) - \widetilde{\theta}(Y)$$

и ожидать, что ее смещение значительно меньше чем было.

Количество выборок m здесь мы делаем достаточно большим, а размеры выборок должны быть те же n, что и у нас на самом деле.

Задача 2. С помощью бутстрэппинга исправить смещение оценки S^2 для выборки размера 20 из $\mathcal{N}(0, 1)$.

3.3.3 Бутстрэп для оценки дисперсии

Использовать метод бутстрэпа можно не только для оценивания смещения, но и, например, для оценки квадратичного смещения. Так мы могли бы взять $\frac{1}{m} \sum_{i=1}^{m} (\widehat{\theta}(Y_{i,1}, \ldots, Y_{i,n}) - \widehat{\theta}(X_1, \ldots, X_n))^2$ в качестве оценки дисперсии $\widehat{\theta}(X_1, \ldots, X_n)$.

Задача 3. Пусть известно $\overline{X}$ из некоторой выборки. Как с его помощью бутстрэппингом оценить дисперсию нашего распределения? Применить метод для $R[0, 1]$.

Этот метод имеет несколько неоспоримых плюсов — он прост в использовании и не требует вычислений, применим даже к весьма громоздким моделям. С другой стороны, мы не можем явным образом оценить его погрешность, а в случае, если оценка $\widehat{\theta}$ значимо промахнулась мимо θ, рискуем неправильно изменить оценку.

Вопрос 6. Предположим, что мы оценили среднее в модели $R[0, \theta]$ с помощью $\overline{X}$. Теперь мы берем новую выборку из $R[0, 2\overline{X}]$ и оцениваем с помощью ее среднего θ. Какую дисперсию будет иметь эта новая оценка?

Получая оценку для дисперсии, мы можем, например, использовать ее для построения асимптотического доверительного интервала на основе асимптотической нормальности. Вместо долгих вычислений в Дельта-методе можно получить оценку бутстрэпом.

Задача 4. Построить асимптотически доверительный интервал для μ/σ, где $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $\mu > 0$ а) с помощью Дельта-метода, б) с помощью оценки дисперсии Бутстрэпом. Сравнить их на выборках размера а) 30 б) 100 в) 1000.

3.3.4 Доверительные интервалы с помощью бутстрэпа

Отсюда мы можем построить доверительный интервал. Методов для этого несколько, рассмотрим наиболее простые.

Рассмотрим оценку $\widehat{\theta}$ для θ. Будем брать выборки из $F_{\widehat{\theta}}$ и строить на основе них оценки $\widehat{\theta}_1, \dots, \widehat{\theta}_m$.

- Normal интервал предлагет брать интервал $(\widehat{\theta} - z_{1-\alpha/2}\widehat{\sigma}, \widehat{\theta} + z_{1-\alpha/2}\widehat{\sigma})$, где $\widehat{\sigma}$ — бутстрэп-оценка для дисперсии $\widehat{\theta}$. Этот интервал работает только для асимптотически нормальных оценок.
- Percentile интервал предлагает взять в качестве интервала для θ диапазон $(\widehat{\theta}_{([\gamma m])}, \widehat{\theta}_{([\beta m])})$ (то есть $[\gamma m]$-ое и $[\beta m]$-ое по возрастанию значения $\widehat{\theta}_i$), где $\beta - \gamma = 1 - \alpha$.
- Pivotal интервал.
  Рассмотрим $\Delta_i = \widehat{\theta}_i - \widehat{\theta}$. Мы ожидаем, что это выборка из величин, близких к $\Delta = \widehat{\theta} - \theta$. Если бы мы знали ф.р. F_Δ, то мы смогли бы построить интервал
  $$\left(F_\Delta^{-1}(\gamma), F_\Delta^{-1}(\beta)\right)$$
  для Δ, где $\beta - \gamma = 1 - \alpha$. Значит мы должны оценить $F_\Delta^{-1}(\gamma)$ и $F_\Delta^{-1}(\beta)$. Заметим, что $F_\Delta(x)$ оценивается величиной
  $$\frac{1}{m}\sum_{i=1}^m I_{\Delta_i \le x}.$$
  Обратная функция к такой функции в точке y — это $[ym]$-я по возрастанию точка из Δ_i. Вывод — упорядочим Δ_i и выберем те из них Δ_-, Δ_+, которые стоят на местах $[\gamma m]$ и $[\beta m]$ по возрастанию. Тогда
  $$(\widehat{\theta} - \Delta_+, \widehat{\theta} - \Delta_-)$$
  и будет нашим интервалом.

  Этот интервал имеет асимптотический уровень доверия $1 - \alpha$.

Задача 5. Сравнить на выборках размера 50 для а) $\mathcal{N}(\theta, 1)$, б) $R[0, \theta]$ доверительные интервалы на основе ОММ, ОМП, бутстрэпа с помощью $\overline{X}$.

Оба интервала легко обобщаются на случай многомерного параметра: мы можем выбрать эллипс, содержащий $(1 - \alpha) \cdot 100\%$ $\widehat{\theta}^*$ или $2\widehat{\theta} - \widehat{\theta}^*$.

В R удобно строить такие выборочные доверительные эллипсы с помощью data.ellipse(x, y, levels=0.9), где x, y задают массивы точек, а levels — уровень.

3.4 Замечания о надежности метода

Несмотря на простую форму, бутстрэп имеет под собой математические основы — такого рода оценки действительно сходятся к оцениваемому параметру для состоятельных оценок $\widehat{\theta}$. Однако, метод имеет скорее прикладное значение — он достаточно прост в использовании даже в сложных моделях.

Бутстрэп опирается на предположение, которое принято формулировать "The population is to the sample as the sample is to the bootstrap samples." Иначе говоря, при генерировании выборок из $F_{\widehat{\theta}}$ они будут отличаться от $\widehat{\theta}$ примерно также как $\widehat{\theta}$ от θ. Если это утверждение не выполнено, то метод может испортить оценку $\widehat{\theta}$.

Percentile интервал опирается на то, что оценка $\widehat{\theta}$ не имеет смещения и ее распределение достаточно симметрично. Из-за этого этот простой метод оказывается достаточно ненадежным — истинное значение параметра будет вылетать из интервала значительно чаще, чем должно (возможно даже никогда в него не попадать). Использовать этот метод стоит только если вы построили график $\widehat{\theta}^* - \widehat{\theta}$ и он оказался симметричным около 0. Pivotal значительно менее придирчив к данным. В следующий раз мы поговорим о более тонких вариантах этого метода.

Задача 6. Построить Percentile и Pivotal интервал для θ, $X_i \sim R[0, \theta]$, на основе смещенной оценки $\max(X_i)$ по 50 наблюдениям. Повторить эксперимент 100 раз. Как часто доверительный интервал накрывал истинное значение параметра?

Для ленивых эти и другие примеры работы метода разобраны в файлах BootstrapConf.R и BootstrapConf.py

4 Ответы на вопросы

Ответ 1. Как выглядит доверительный интервал на основе $\max X_i$, $X_i \sim R[0,\theta]$?

Величина $\max X_i$ имеет функцию распределения

$$F_{\max X_i}(x) = \mathbf{P}(\max X_i \le x) = \mathbf{P}(X_1 \le x, \ldots, X_n \le x) = F_{X_1}(x)^n = \frac{x^n}{\theta^n},\ x \in [0,\theta].$$

Значит

$$F_{\max X_i/\theta}(x) = x^n,\ x \in [0,\theta].$$

Выбирая произвольное $\beta \in [0,\alpha]$, мы получаем

$$\mathbf{P}\left(\frac{\max X_i}{\theta} \in [x_\beta, x_{1-\alpha+\beta}]\right) = 1-\alpha,$$

где x_β — β-квантиль ф.р. x^n. Решая уравнение $F(x) = \beta$, получаем, что

$$\mathbf{P}\left(\frac{\max X_i}{\theta} \in [\sqrt[n]{\beta},\, \sqrt[n]{1-\alpha+\beta}]\right) = 1-\alpha,$$

откуда доверительным интервалом уровня доверия $1-\alpha$ будет любой интервал вида

$$\left(\frac{\max X_i}{\sqrt[n]{1-\alpha+\beta}}, \frac{\max X_i}{\sqrt[n]{\beta}}\right).$$

В принципе β можно брать любым, но выгоднее всего брать $\beta = \alpha$, тогда интервал будет кратчайшей длины.

Ответ 2. Показать, что для $X_i \sim Poiss(\lambda)$

$$(\widehat{\theta}_1, \widehat{\theta}_2), \quad \widehat{\theta}_1 = \frac{1}{2n} y_{1-\alpha/2, 2\sum_{i=1}^n X_i}, \quad \widehat{\theta}_2 = \frac{1}{2n} y_{\alpha/2, 2\sum_{i=1}^n X_i + 2},$$

является доверительным интервалом для λ.

Пусть $X_i \sim Poiss(\lambda)$. Тогда $X_1 + \ldots + X_n \sim Poiss(\lambda n)$. Поскольку при любом целом x

$$F_{Poiss(\lambda n)}(x) = \mathbf{P}(N_1 \le x) = \mathbf{P}(\tau_1 + \cdots + \tau_{x+1} \ge 1) = 1 - F_{\Gamma(x+1,\lambda n)}(1),$$

где N_t — пуассоновский процесс с параметром λn, τ_i — моменты его скачков. Я использую пуассоновский процесс, чтобы показать связь между пуассоновским и экспоненциальным распределениями, но можно сделать это и напрямую без высоких материй, важно здесь то, что ф.р. пуассоновской величины тесно связана с ф.р. гаммы. При этом

$$F_{\Gamma(x+1,n\lambda)}(1) = F_{\chi^2_{2x+2}}(2n\lambda).$$

Значит,

$$\mathbf{P}(X_1 + \cdots + X_n \le x) = 1 - F_{\chi^2_{2x+2}}(2n\lambda).$$

Отсюда видно, что ф.р. пуассоновского распределения монотонно убывает по λ.

Аналогично

$$\mathbf{P}(X_1 + \cdots + X_n \ge x) = F_{\chi^2_{2x}}(2n\lambda).$$

Значит, соотношения

$$\mathbf{P}_{\widehat{\theta}_1}(T \le x) = \alpha/2, \quad \mathbf{P}_{\widehat{\theta}_2}(T \ge x) = \alpha/2$$

равносильны

$$\widehat{\theta}_1 = \frac{1}{2n} y_{1-\alpha/2, 2x}, \quad \widehat{\theta}_2 = \frac{1}{2n} y_{\alpha/2, 2x+2},$$

где $y_{\beta,z}$ — квантиль χ^2_{2z} уровня β. Получаем требуемый интервал.

Ответ 3. Почему асимптотический доверительный интервал для асимптотически нормальной оценки с непрерывной дисперсией имеет вид

$$\left(\widehat{\theta}(X_1,\ldots,X_n) - \frac{z_{1-\alpha/2}\sigma(\widehat{\theta}(X_1,\ldots,X_n))}{\sqrt{n}}, \widehat{\theta}(X_1,\ldots,X_n) + \frac{z_{1-\alpha/2}\sigma(\widehat{\theta}(X_1,\ldots,X_n))}{\sqrt{n}}\right)?$$

В силу асимптотической нормальности

$$\sqrt{n}\frac{\widehat{\theta}(X_1,\ldots,X_n)-\theta}{\sigma(\theta)} \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

Заметим, что $\widehat{\theta}$ состоятельна для θ (в силу асимптотической нормальности), а σ — непрерывная функция, то $\sigma(\widehat{\theta}(X_1,\ldots,X_n))$ состоятельная оценка $\sigma(\theta)$.

Значит, при $n \to \infty$

$$\sqrt{n}\frac{\widehat{\theta}(X_1,\ldots,X_n)-\theta}{\sigma(\theta)} = \sqrt{n}\frac{\widehat{\theta}(X_1,\ldots,X_n)-\theta}{\sigma(\theta)} \cdot \frac{\sigma(\theta)}{\sigma(\widehat{\theta}(X_1,\ldots,X_n))} \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

Здесь мы воспользовались тем, что если одна последовательность сходится по распределению к Z, а вторая к константе c, то их произведение сходится по распределению к cZ. Отсюда при любом $\beta \leq \alpha$

$$\mathbf{P}\left(n^{-1/2}z_\beta/\sigma(\widehat{\theta}(X_1,\ldots,X_n)) \leq \widehat{\theta}(X_1,\ldots,X_n)-\theta \leq n^{-1/2}z_{1-\alpha+\beta}/\sigma(\widehat{\theta}(X_1,\ldots,X_n))\right) \to 1-\alpha. \qquad (2)$$

Отсюда приходим к требуемой форме интервала.

Ответ 4. Почему для нормального распределения имеет смысл брать симметричный интервал с квантилями уровня $\alpha/2$ и $1-\alpha/2$?

Покажем, что в (2) выбирая $\beta = \alpha/2$, мы получим кратчайший интервал. Действительно, длина интервала пропорциональна величине $z_{1-\alpha+\beta} - z_\beta$. Максимизируем эту величину. Дифференцируя по α, получаем

$$z'_{1-\alpha+\beta} - z'_\beta = \frac{1}{\Phi^{-1}(z_{1-\alpha+\beta})} - \frac{1}{\Phi^{-1}(z_\beta)} = \frac{1}{\varphi(z_{1-\alpha+\beta})} - \frac{1}{\varphi(z_\beta)},$$

где Φ, φ — ф.р. и плотность $\mathcal{N}(0,1)$. Здесь мы воспользовались формулой производной обратной функции.

Плотность φ монотонно убывает при положительных аргументах и монотонно возрастает при отрицательных, откуда ноль производной будет давать нам минимум. При этом функция φ каждое значение принимает дважды — в точке u и $-u$. Значит

$$z_{1-\alpha+\beta} = -z_\beta, \quad \Phi(z_{1-\alpha+\beta}) = 1-\Phi(z_\beta),$$

откуда и вытекает соотношение $\beta = \alpha/2$.

Ответ 5. Показать для многомерной нормальной выборки $(X_1,\ldots,X_k) \sim \mathcal{N}(\vec{\mu},\Sigma)$, что она попадает в эллипсоид с центром $\vec{\mu}$

$$(\vec{X}-\vec{\mu})^{\mathrm{T}}\Sigma^{-1}(\vec{X}-\vec{\mu}) \leq x$$

с вероятностью $F_{\chi^2_k}(x)$.

Поскольку Σ положительно определена, то $\Sigma = S^{\mathrm{T}}S$ для некоторой S. При этом вектор $S^{-1}(\vec{X}-\vec{\mu})$ имеет распределение

$$\mathcal{N}\left(\vec{0}, S^{-1}\Sigma(S^{-1})^{\mathrm{T}}\right) = \mathcal{N}\left(\vec{0}, E\right).$$

Остается заметить, что тогда

$$(S^{-1}(\vec{X}-\vec{\mu}))^{\mathrm{T}}(S^{-1}(\vec{X}-\vec{\mu})) = (\vec{X}-\vec{\mu})^{\mathrm{T}}\Sigma^{-1}(\vec{X}-\vec{\mu}) \sim \chi^2_k.$$

Ответ 6. Предположим, что мы оценили среднее в модели $R[0,\theta]$ с помощью $\overline{X}$. Теперь мы берем новую выборку из $R[0,2\overline{X}]$ и оцениваем с помощью ее среднего θ. Какую дисперсию будет иметь эта новая

оценка?
Величины $Y_i \sim R[0, 2\overline{X}]$ могут быть представлены как $Y_i = 2\overline{X}R_i$, где $R_i \sim R[0,1]$ независимы от $\overline{X}$. Тогда

$$\mathbf{E}Y_i^2 = 4\mathbf{E}\overline{X}^2\mathbf{E}R_i^2 = 4(\mathbf{D}\overline{X} + (\mathbf{E}\overline{X})^2)(\mathbf{D}R_i + (\mathbf{E}R_i)^2) = 4\left(\frac{\theta^2}{12} + \frac{\theta^2}{4}\right)\left(\frac{1}{12} + \frac{1}{4}\right) = \frac{4}{9}\theta^2.$$

При этом $\mathbf{E}Y_i = 2\mathbf{E}\overline{X}\mathbf{E}R_i = \theta/2$. Следовательно, дисперсия этой оценки $\frac{7}{36}\theta^2$, то есть больше той дисперсии $\frac{1}{12}\theta^2$, которая была бы у $\overline{X}$. Это логично, дополнительная рандомизация не уменьшает дисперсию.