СЕМ, Л-14
(2021-01-14)

Доверителни интервали

Постановка.

- Случайна величина X
- $F_X(x, \theta)$ е разпределение, което искаме да разберем, като знаем че то зависи параметрично от някакъв параметър θ
- $\theta = (\theta_1, \ldots, \theta_s)$ са s на брой параметри (например при нормалното разпределение са $s = 2$: средно μ и стандартно отклонение σ)

$\overrightarrow{X}(X_1, \ldots, X_n)$ е вектор от n независими еднакво разпределени наблюдения над X (прототипи на X). На база на тези наблюдения, които в крайна сметка ще бъдат сведени до някакви числа (за модела/за експеримента) трябва да намерим някаква оценка $\underbrace{\hat{\theta} = \hat{\theta}(\overrightarrow{X})}_{\text{точкова оценка}}$, която да я вземем близо до θ, така че да определи това разпределение $F_X(x, \theta)$.

Проблемът с точковата оценка е, че тя варира в зависимост от конкретната извадка. Това е очаквано, тъй като различните извадки дават различни резултати. Затова, освен точечната оценка $\hat{\theta}$, би било много по-информативно да разполагаме с интервал от стойности, в който с дадена вероятност се намира истинският параметър θ. Такъв интервал се нарича *доверителен интервал*.

Ще разглеждаме само едномерни параметри θ $(\theta \in \mathbb{R})$.

Цел. Ще търсим две числа L и U, такива, че:

$$\mathbb{P}(L \leq \theta \leq U) = 1 - \alpha,$$

където α обикновено е число между 0 и 1 $(0 < \alpha < 1)$ и има следния смисъл: колкото по-малко е α, толкова по-широки интервали ще се получават, за да може с по-голяма вероятност (повече „сигурност“) да покрием истинския параметър θ. Стандартно, $\alpha = 0.05$ (което отговаря на 95% доверителен интервал). Стойност като $\alpha = 0.10$ се използва за по-маловажни изследвания или предварителни анализи, докато за медицински или много критични цели може да се използва $\alpha = 0.01$.

Дефиниция (Централна статистика ЦС). Казваме, че $T = T(\overrightarrow{X}, \theta)$ е централна статистика, ако:

1) T е монотонна по θ
2) $\mathbb{P}(T < x) = F_T(x)$ не зависи от θ (T е функция на θ, но разпределението и не зависи от θ)

⊕ Имаме $\overrightarrow{X}$ (вектор от наблюдения) и искаме да намерим някакъв доверителен интервал: (L, U) за θ, тоест $\theta \in (L, U)$. Целта ни е да имаме някакво ниво на доверие $1 - \alpha = \mathbb{P}(q_1 < T < q_2)$. За улеснение ще допуснем, че T расте по θ (но тя може и да намалява по θ).

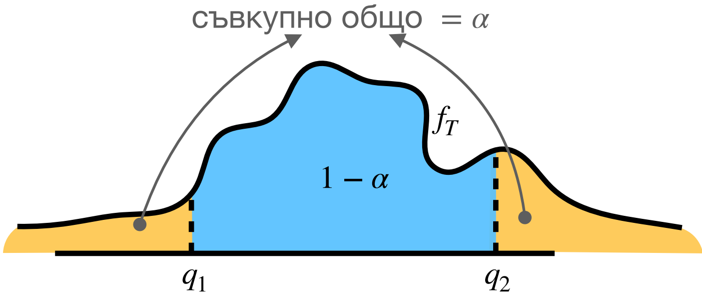
Тъй като T е монотонна по θ, то:

$$1 - \alpha = \mathbb{P}(q_1 < T < q_2) \overset{\substack{\text{при фиксиран}\\ \text{вектор на}\\ \text{наблюдения } \overrightarrow{X}}}{=} \mathbb{P}\left(T^{-1}(q_1) < \theta < T^{-1}(q_2)\right),$$

тъй като сме допуснали, че T расте по θ. Ако T намаляваше по θ, щяхме да имаме $1 - \alpha = \mathbb{P}\left(T^{-1}(q_2) < \theta < T^{-1}(q_1)\right)$. Следователно интервалът ще е:

$$L = T^{-1}(q_1), U = T^{-1}(q_2).$$

f_T - плътността на централната статистика (ЦС)

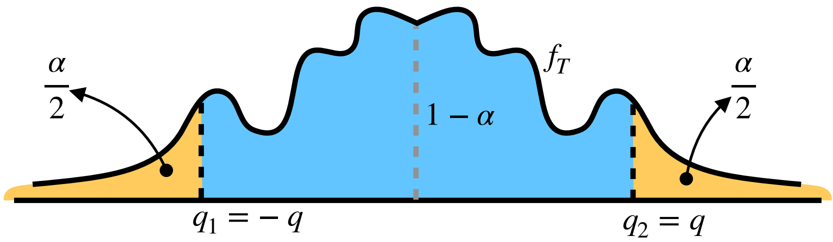


Има много начини, по които може да изберем q_1 и q_2, така че вероятността между тях да е $1 - \alpha$.

Имаме параметър $\theta \in \mathbb{R}$, който искаме да оценим. Търсим L и U, които да зависят от наблюденията $\overrightarrow{X}$ и $L(\overrightarrow{X}) < U(\overrightarrow{X})$. Те образуват така наречения *доверителен интервал* за θ с ниво на доверие $1 - \alpha = \mathbb{P}\left(\theta \in (L, U)\right)$.

T е централна статистика за θ, ако удовлетворява дефиницията за ЦС.

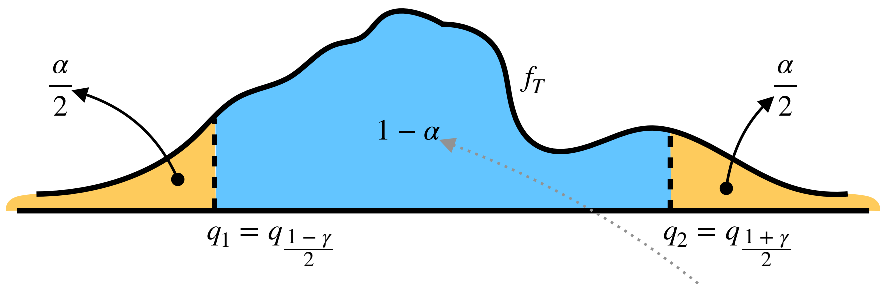
⊕ Ако T е симетрична случайна величина,



тогава се търсят такова q_1 и q_2, че $q_1 = -q = -q_2$, за които

$\mathbb{P}(T < -q) = 1 - \mathbb{P}(T < q) = \dfrac{\alpha}{2}$ (за да имаме по средата вероятност $1 - \alpha$)

По-общо, ако имаме някакво несиметрично разпределение на T:



проверка : $\mathbb{P}(T < q_2) - \mathbb{P}(T < q_1) = \dfrac{\alpha}{2} + 1 - \alpha - \dfrac{\alpha}{2} = 1 - \alpha$

Има много възможни начини за избор на границите q_1 и q_2 на доверителния интервал. Въпреки това, формулите, представени по-горе, не са произволни. Те са стандартизирани и добре установени методи в статистическата практика, които произлизат директно от свойствата на пивотните величини.

$$\underbrace{1-\alpha}_{\text{фиксирана}} = \mathbb{P}(q_1 < T < q_2) = \mathbb{P}\big(\underbrace{T^{-1}(q_1) < \theta < T^{-1}(q_2)}_{T \text{ нараства}}\big).$$

Тази вероятност $1 - \alpha$ е фиксирана и се задава предварително от изследователите. Какво може да оптимизираме ние като математици? Може да търсим q_1 и q_2 такива, за които е изпълнено:

$$\min_{\substack{q_1 < q_2 \\ 1-\alpha = \mathbb{P}(q_1 < T < q_2)}} = \Big\{ \underbrace{\left| T^{-1}(q_2) - T^{-1}(q_1) \right|}_{\substack{\text{искаме интервал с минимална} \\ \text{дължина, за да ограничим} \\ \text{възможните стойности на} \\ \text{параметъра } \theta \text{ максимално}}} \Big\}.$$

Тоест минимизираме доверителния интервал при фиксирано ниво на доверие!

$\oplus$ $\underbrace{X \in \mathcal{N}(\mu, \sigma^2)}_{\substack{\text{случайна величина, която} \\ \text{искаме да изучаваме}}}$, σ^2 е известно,

тоест интересуваме се само от параметъра $\mu = \theta$ (едномерен). Искаме да видим как може да оценим μ и да намерим за него доверителен интервал.

Ние знаем, че $\hat{\mu} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Тогава $T(\overrightarrow{X}, \mu) = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$, σ е известно число. Това е така, защото:

$$\sum_{i=1}^{n} X_i \overset{\substack{\text{линейност} \\ \text{на } \mathcal{N} \text{ и незав.}}}{\sim} \mathcal{N}(n\mu, n\sigma^2) \Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}\right) = \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

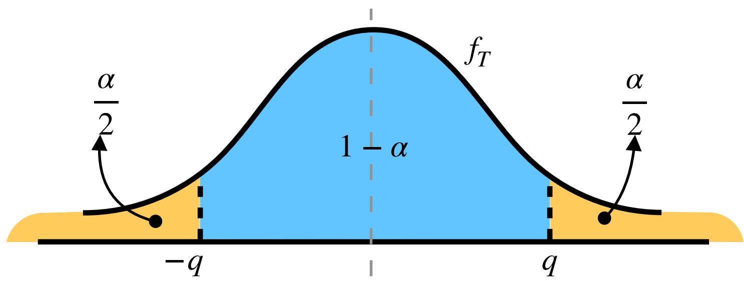
По-подробно обяснение на свойството линейност на нормалното разпределение:

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \sim \mathcal{N}(x, y^2);\ \mathbb{E}\overline{X}_n = \frac{n\mu}{n} = x;\ \mathrm{Var}[\overline{X}_n] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n} = y^2$$

T е намаляваща функция по μ и $\mathbb{P}(T \le x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{y^2}{2}}\,\mathrm{d}\,y$ - не зависи от $\mu \overset{def.}{\Rightarrow}$

T е централна статистика за μ (тя е монотонна и намаляваща по μ и нейното разпределение съвпада с $\mathcal{N}(0,1)$, тоест не зависи от θ)

Тогава, $1 - \alpha = \mathbb{P}(-q < T < q)$, тъй като $\mathcal{N}(0,1)$ е симетрично:



Това ни гарантира, че $(-q, q)$ ще е най-късия интервал, тъй като в $(-\infty, -q)$ и (q, ∞) е сбита най-малко маса (навсякъде извън тези интервали е сбита повече маса)

$q = q_{\frac{1}{2}+\frac{1-\alpha}{2}}$, който квантил го има в таблицата за нормалното стандартно разпределение.

$$1-\alpha=\mathbb{P}(-q<T<q)=\mathbb{P}\left(-q_{\frac{1}{2}+\frac{1-\alpha}{2}}<\frac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}}<q_{\frac{1}{2}+\frac{1-\alpha}{2}}\right)\Rightarrow$$

$$\Rightarrow\mathbb{P}\left(\mu\in(\underbrace{\overline{X}_n-\frac{\sigma}{\sqrt{n}}\cdot q_{\frac{1}{2}+\frac{1-\alpha}{2}}}_{L},\underbrace{\overline{X}_n+\frac{\sigma}{\sqrt{n}}\cdot q_{\frac{1}{2}+\frac{1-\alpha}{2}}}_{U})\right)=1-\alpha\Rightarrow$$

$$\Rightarrow L=\overline{X}_n-\frac{\sigma}{\sqrt{n}}\cdot q_{\frac{1}{2}+\frac{1-\alpha}{2}};\ U=\overline{X}_n+\frac{\sigma}{\sqrt{n}}\cdot q_{\frac{1}{2}+\frac{1-\alpha}{2}}$$

$\oplus$ $X\sim\mathcal{N}(\mu,\sigma^2)$, но този път не знаем σ.

Как да конструираме доверителен интервал само за μ? Припомняме, че $\underbrace{\hat{\mu}=\overline{X}_n}_{\text{независимо дали знаем или не } \sigma}$ и оценката за дисперсията е $s^2=\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$ и е неизместена.

Фактора $\dfrac{1}{n-1}$ го има, тъй като тя е неизместена оценка за дисперсията.

Твърдение. Имаме, че $X\sim\mathcal{N}(\mu,\sigma^2)$ и $\overrightarrow{X}=(X_1,\ldots,X_n)$ са наблюдения над X. Тогава е вярно, че:

- $\hat{\mu}$ е независимо от s^2 : $\hat{\mu}\perp\!\!\!\perp s^2$
- $(n-1)\dfrac{s^2}{\sigma^2}\sim\mathcal{X}^2(n-1)$.

$T=\dfrac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}}$. Ако знаехме σ, последното щеше да е разпределено като $\mathcal{N}(0,1)$ неизвестно (както направихме в предходния пример), но ние не знаем σ. Но друго, което знаем е, че:

$$T=\frac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}}=\frac{\overset{\sim\mathcal{N}(0,1)}{\frac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}}}}{\sqrt{\frac{n-1}{n-1}\cdot\underset{\sim\mathcal{X}^2(n+1)}{\frac{s^2}{\sigma^2}}}}=\frac{\overline{X}_n-\mu}{\frac{s}{\sqrt{n}}}\text{ е централна статистика!}$$

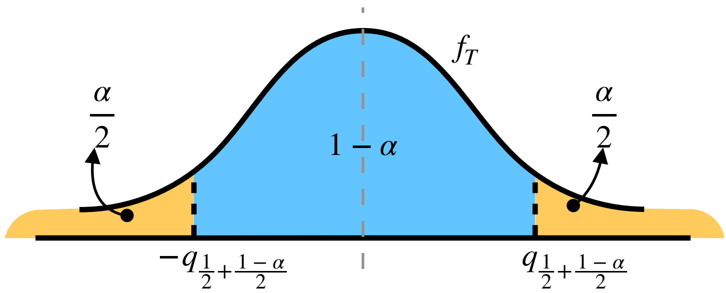
Тъй като $\dfrac{\overline{X}_n-\mu}{\frac{s}{\sqrt{n}}}=\dfrac{Z}{\sqrt{\frac{Y}{n-1}}}$, където $Z\sim\mathcal{N}(0,1)$, $Y\sim\mathcal{X}^2(n+1)$ и $Z\perp\!\!\!\perp Y$.

Заключения.

- T е намаляваща по μ
- $T \in t(n-1)$ и не зависи от $\mu \Rightarrow T$ е централна статистика за μ.

$$T = \frac{Z}{\sqrt{\frac{Y}{n-1}}};\ Z = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}};\ Y = \frac{(n-1)s^2}{\sigma^2}.$$

T е симетрична, тъй като в числителя имаме симетрична случайна величина $(Z) \Rightarrow$



$$\Rightarrow \mathbb{P}\left(-q_{\frac{1}{2}+\frac{1-\alpha}{2}} < T < q_{\frac{1}{2}+\frac{1-\alpha}{2}}\right) = \mathbb{P}\left(\mu \in \left(\overline{X}_n - q_{\frac{1}{2}+\frac{1-\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \overline{X}_n + q_{\frac{1}{2}+\frac{1-\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right)\right)$$

тоест имаме, че L и U зависят от s и квантилите не са от нормалното разпределение а от t — Student's разпределението.

Бележки. $T - \frac{\overline{X}_n - \mu}{\frac{s}{\sqrt{n}}} = \frac{Z}{\sqrt{\frac{Y}{n-1}}}$, $Z \perp\!\!\!\perp Y$, $Z \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{X}^2(n-1)$

- $Y = \sum_{j=i}^{n-1} V_i$, където $\{V_i\}_{i=1}^{n-1}$ са независими и еднакво разпределени (i.i.d.) с $\mathcal{X}^2(1)$.

  От ЗГЧ: $\frac{Y}{n-1} \xrightarrow[n\to\infty]{\text{п.с.}} \mathbb{E}[V_1] = 1$. Следователно за големи n : $T \approx \frac{Z}{1} \approx \mathcal{N}(0,1)$;

- Ако знаем дисперсията: $\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)$ за произволни $X_1, \ldots, X_n, \ldots$

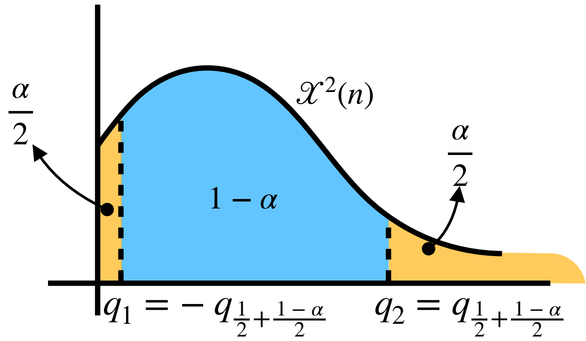
  $\mathbb{E}[X_1] = \mu;\ \mathrm{Var}[X_1] = \sigma^2$.

$\oplus$ $X \sim \mathcal{N}(\mu, \sigma^2)$, знаем μ, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$. Искаме да оценим дисперсията.

Трябва да конструираме централна статистика. ЦС не трябва да зависи от σ, но в $\hat{\sigma}^2$, X_j зависи от σ и трябва да отстраним тази зависимост.

Нагаждаме: $\dfrac{n\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{n} \Bigg(\underbrace{\dfrac{X_i - \mu}{\sigma}}_{Z_i^2}\Bigg) = \sum_{i=1}^{n} Z_i^2$, но $Z_i \sim \mathcal{N}(0,1)$, тъй като центрираме и нормираме с неизвестна дисперсия. Следователно $\sum_{i=1}^{n} Z_i^2 \sim \mathcal{X}^2(n)$, защото сумира n квадрати на независими нормални стандартно разпределени $Z \Rightarrow$ статистиката $T = \dfrac{n\hat{\sigma}^2}{\sigma^2}$ е централна по дефиниция (разпределението и е $\mathcal{X}^2(n)$ – не зависи от σ^2 и тя е монотонно намаляваща по σ).

За квантилите:

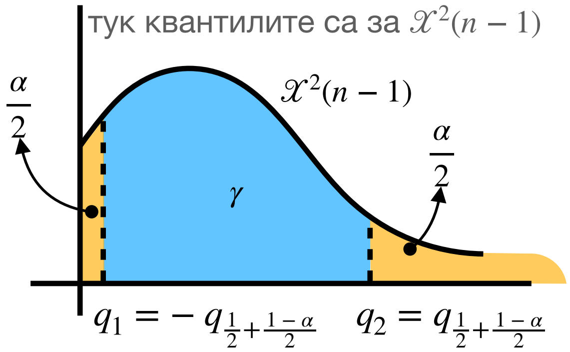


$$1-\alpha = \mathbb{P}(q_1 < T < q_2) = \mathbb{P}\left(\frac{n\hat{\sigma}^2}{q_2} < \frac{n\hat{\sigma}^2}{q_1}\right) \Rightarrow \begin{cases} L = \frac{n\hat{\sigma}^2}{q_2} \\ U = \frac{n\hat{\sigma}^2}{q_1} \end{cases}$$

$\oplus$ $X \sim \mathcal{N}(\mu, \sigma^2)$, не знаем μ и се интересуваме от σ^2. Знаем, че s^2 е неизместена оценка за σ^2:

$(n-1)s^2 = \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$ и нашата цел е да конструираме статистика за σ^2, тоест да намерим ЦС, отговаряща на дефиницията. Ако разделим на σ^2, от дясно не можем да направим никакво заключение: $\dfrac{(n-1)s^2}{\sigma^2} = \dfrac{\sum_{i=1}^{n} (X_i - \overline{X}_n)^2}{\sigma^2}$, но знаем от твърдението, че $T = \dfrac{(n-1)s^2}{\sigma^2} \sim \mathcal{X}^2(n-1)$ (с една степен по-малко, защото сме изхабили една степен за оценката на μ). От n наблюдения все едно имаме $n-1$ наблюдения.

$\Rightarrow$ T е централна статистика (монотонна намаляваща по σ^2 и независима от σ^2)



$$1-\alpha = \mathbb{P}\left(q_1 < \frac{(n-1)s^2}{\sigma^2} < q_2\right) = \mathbb{P}\left(\frac{(n-1)s^2}{q_2} < \sigma^2 < \frac{(n-1)s^2}{q_1}\right)$$

$$L = \frac{1}{q_2}(n-1)s^2;\ U = \frac{1}{q_1}(n-1)s^2$$

При голямо n $(n > 30$ например) може да ползваме както от предходния пример - все едно знаем σ.

Проверка на хипотези

Теорията за проверка на хипотези не е откритие на един човек, а се развива през 20-ти век чрез приносите на няколко ключови статистици. В основата на съвременната теория стоят Роналд Фишър (1890 – 1962, Великобритания), който въвежда основите, и Йерзи Сплава-Нейман (1894 – 1981, Поляк) с Егон Шарп Пирсън (1895 – 1980, Великобритания), които я формализират. В по-ранни времена зачатъците на идеите могат да се открият дори в работите на Джон Арбътнот през 18-ти век. Тя най-често се задава със следната математическа постановка:

X е случайна величина с функция на разпределение $F_X(x,\theta)$, нулева хипотеза H_0 и алтернативна такава H_1, където θ е оценяваният параметър.

- $H_0 : \theta = \theta_0$
- $H_1 : \theta = \theta_1$

Искаме да конструираме някакво множество $W \in \mathbb{R}^n$ такова, че ако $\overrightarrow{X}$ попадне в W $(\overrightarrow{X} \in W)$, тогава отхвърляме H_0 (в полза на H_1). Ако векторът от наблюдения $\overrightarrow{X}$ не попадне в W, то тогава не отхвърляме H_0.

H_0 и H_1 се наричат прости хипотези (нулева и алтернативна). Прости хипотези са $\theta = \theta_1$ (число). Сложни хипотези са $\theta > \theta_i$, $\theta \neq \theta_i$, $\theta \in I$ и т.н.

Цел. При наблюденията $\overrightarrow{X} \in \mathbb{R}^n$, търсим да конструираме $W \subseteq \mathbb{R}^n$: ако $\overrightarrow{X} \in W$, то отхвърляме H_0 и приемаме H_1, а ако $\overrightarrow{X} \in \overline{W}$, то приемаме H_0.

$$W \subseteq \mathbb{R}^n : \begin{cases} \overrightarrow{X} \in W \Rightarrow \text{отхвърляме } H_0 \\ \overrightarrow{X} \in \overline{W} \Rightarrow \text{приемаме } H_0 \end{cases}$$

Грешки, които може да допуснем:

- Грешка от I$^{\text{-ви}}$ род: Да отхвърлим H_0, когато H_0 е вярна, тоест:

$$\alpha = \mathbb{P}(\overrightarrow{X} \in W \,|\, H_0)$$

- Грешка от II$^{\text{-ри}}$ род: При положение, че е вярна хипотезата H_1, ние сме приели H_0, тоест

$$\beta = \mathbb{P}(\overrightarrow{X} \in \overline{W} \,|\, H_1)$$

$$\pi = 1 - \beta \text{ се нарича } \textit{мощност} \text{ на } W$$

$\oplus$ H_0 : дадена ваксина е вредна $(\theta = \theta_0)$. H_1 : ваксината не е вредна $(\theta = \theta_1)$

Има по-голям резон за H_0 да вземем по-опасната/рисковата хипотеза. Това е тази хипотеза, която е по-вероятно да я отхвърлим. Това е така, защото грешката от I$^{\text{-ви}}$ род ще бъде контролирана/задавана от изследователя (той ще казва дали иска/допуска да е 0.01, 0.05, 0.10 и т.н.)

Дефиниция (Оптимална критична област). При фиксирана грешка от I$^{\text{-ви}}$ род α, $W^* \subseteq \mathbb{R}^n$ се нарича оптимална критична област (ОКО), ако

$$\mathbb{P}(\overrightarrow{X} \notin W^* \,|\, H_1) = \min_{\substack{\underbrace{W \subseteq \mathbb{R}^2}_{\text{всички критични области}} \\ \alpha = \mathbb{P}(X \in W \,|\, H_0)}} \mathbb{P}(\overrightarrow{X} \notin W \,|\, H_1).$$

Постановка. X е случайна величина; $F_X(x, \theta)$ е разпределението на X, което зависи от накакъв параметър θ, но допускаме, че $f_X(x, \theta)$ е плътността на X (тоест допускаме, че $\dfrac{\partial x}{\partial} F_X(x, \theta)$ съществува). Въвеждаме:

$$f_{\overrightarrow{X}}(x, \theta) = \underbrace{L(X, \theta)}_{\text{функция на правдоподобие}} = \prod_{i=1}^{n} f_X(x_i, \theta), \text{ където } x \in \mathbb{R}^n \text{ и } x = (x_1, \ldots, x_n).$$

Тогава е верен следния резултат:

Лема (Нейман-Пирсън). Нека X удовлетворява горните условия от постановката и нека тестваме следната хипотеза:

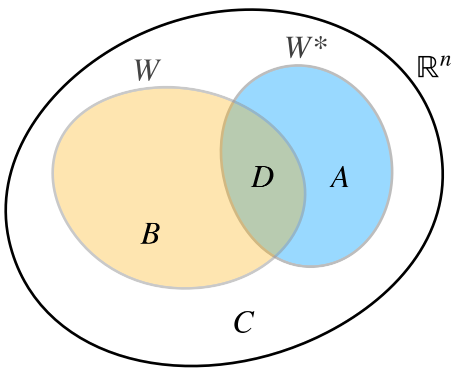
$H_0 : \theta = \theta_0$ срещу $H_1 : \theta = \theta_1$.

Ако $L_0(x) = L(x, \theta_0)$ и $L_1(x) = L(x, \theta_1)$ и

$\exists k \geq 0 : W^* \subseteq \{x \in \mathbb{R}^n : L_1(x) \geq kL_0(x)\}, \quad \overline{W}^* \subseteq \{X \in \mathbb{R}^2 : L_1(x) \leq kL_0(x)\}$ и $\alpha = \mathbb{P}(X \in W^* \,|\, H_0)$ е зададена, то W^* е ОКО (оптимална критична област).

Доказателство:

$$\underbrace{\alpha = \mathbb{P}(\overrightarrow{X} \in W^* \,|\, H_0) = \mathbb{P}(\overrightarrow{X} \in W \,|\, H_0)}_{\text{имаме}} \Rightarrow \underbrace{\mathbb{P}(\overrightarrow{X} \notin W^* \,|\, H_1) \leq \mathbb{P}(\overrightarrow{X} \notin W \,|\, H_1)}_{\text{искаме да докажем}}$$



$$\mathbb{P}(\overrightarrow{X} \notin W \,|\, H_1) \overset{\theta=\theta_1}{=} \int_{\overline{W}} L_1(x)\mathrm{d}\,x$$

$$= \int_A L_1(x)\mathrm{d}\,x + \int_C L_1(x)\mathrm{d}\,x + \underbrace{\int_B L_1(x)\mathrm{d}\,x - \int_B L_1(x)\mathrm{d}\,x}_{} =$$

$$= \int_{\overline{W}^*} L_1(x)\mathrm{d}\,x + \int_A L_1(x)\mathrm{d}\,x - \int_B L_1(x)\mathrm{d}\,x$$

$$= \mathbb{P}(\overrightarrow{X} \in W^* \,|\, H_1) + \underbrace{\int_A L_1(x)\mathrm{d}\,x - \int_B L_1(x)\mathrm{dx}}_{\overset{?}{\geq} 0}$$

$\geq \mathbb{P}(\overrightarrow{X} \notin W^* \,|\, H_1)$, което искахме да докажем.

Тоест всичко се свежда до това да проверим, че $\int_A L_1(x)\mathrm{d}\,x - \int_B L_1(x)\mathrm{d}\,x \geq 0$, но

$$\int_A L_1(x)\mathrm{d}\,x - \int_B L_1(x)\mathrm{d}\,x \geq \underbrace{k\int_A L_0(x)\mathrm{d}\,x - k\int_B L_0(x)\mathrm{d}\,x}_{\stackrel{?}{=}0}$$

$$\begin{aligned}\alpha = \int_{W*} L_0(x)\mathrm{d}\,x &= \int_A L_0(x)\mathrm{d}\,x + \cancel{\int_D l_0(x)\mathrm{d}\,x} \\ = \int_W L_0(x)\mathrm{d}\,x &= \int_B L_0(x)\mathrm{d}\,x + \cancel{\int_D L_0(x)\mathrm{d}\,x} \\ \Rightarrow \int_A L_0(x)\mathrm{d}\,x &= \int_B L_0(x)\mathrm{d}\,x\,.\end{aligned}$$