Математическая статистика

1. Теория статистического оценивания неизвестных параметров распределения
    1. ММП
    2. ММ
    3. Байевское оценивание
    4. Доверительные интервалы
2. Проверка статистических гипотез о параметрах модели, о виде распределения

Генеральная совокупность, выборука, статистическая оценка

$\{\Omega, \mathbb{A}, P\}$ = "случайная величина" = "закон распределения" = "генеральная совокупность"

<u>Опр</u> Генеральной совокупностью называется совокупность всех мыслимых наблюдений, которые могла быть получены при данном реальном комплексе условий.

<u>Опр</u> Выборка из данной генеральной совокупности ξ - это результаты ограниченного ряда наблюдений случайной величины ξ. Выборка - это эмпирический аналог генеральной совокупности

$\vec{x_n} = \{x_1, x_2, \ldots, x_n\}$
x_i - i-ое наблюдение/измерение случайной величины ξ

Св-ва выборки:

1. x_i - случайная величина
2. условия стохастического эксперимента не меня.тся от наблюдения к наблюдению
3. все наблюдения в выборке взаимно независимы

$\implies F_{x_i}(z) = P(x_i < z) = P(\xi < z) = F_\xi(z)$

$F_{\vec{x_n}}(z) = P(x_1 < z, \ldots, x_n < z) = \Pi_{i=1}^n P(x_i < z_i) = \Pi_{i=1}^n P(x_i < z_i) = \Pi_{i=1}^n P(x_i < z_i) = \Pi_{i=1}^n F_{x_i}(z_i$

Статистическое оценивание параметров

<u>Опр</u> $\forall$ функции $f(\vec{x_n})$ называется статистикой.

<u>Опр</u> Статистическая оценка - это статистика
$\hat{\theta}(\vec{x_n})$, использующая в качестве неизвестного значения параметра θ
1. $\hat{\theta}(\vec{x_n})$ - это случайная величина
2. $\hat{\theta}(\vec{x_n})$ не зависит от θ

1. Эмпирическая функция распределения

$$\hat{F}_n(z) = \frac{\nu(z)}{n} = \frac{1}{n}\sum_{i=1}^{n} I(x_i < z)$$

$$I(x_i < z) = \begin{cases} 1, x_i < z \\ 0, x_i \geq z \end{cases}$$

$\nu(z)$ - число компонентов в выборке $< z$ $(x_i < z)$

2. Эмпирическая плотность распределения

$$\hat{p_n}(z) = \frac{\nu_k}{n\Delta_k(z)}$$

$k(z)$ - порядковый номер интервала группирования, который $\ni z$
$\Delta_k(z)$ - ширина интервала
$\nu_k(z)$ - число наблюдений $\in$ в $k(z)$ - ый интервал группирования

$$p(x) = \frac{dF(x)}{dx}$$

Функция Стерджесса: $k = k(n) = 1 + [\log_2(n)]$ (предполагает кол-во интервалов, которое следует выбрать)

Характеристика по определению	**Её возможная оценка**
$\nu_k = M\xi^k$ - начальный момент k-го порядка, $a = M\xi$	$\hat{\nu_k}(\vec{x_n}) = \frac{1}{n}\sum_{i=1}^{n} x_i^k$, $\bar{x} = \frac{1}{n} sum_{i=1}^{n} x_i$ - выборочное среднее
$\mu_k = M(\xi - M\xi)^k$ - центральный момент k -го порядка, $D\xi = \mu_2 = M(\xi - M\xi)^2$	$\hat{\mu_k}(\vec{x_n}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$, $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, $S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Свойства надежных статистических оценок:

1. Состоятельность
2. Несмещенность
3. Эффективность

1. Состоятельность

Опр. Оценка $\hat{\theta}(\vec{x_n})$ для параметра θ называется состоятельной, если $\hat{\theta}(\vec{x_n}) \longrightarrow \theta$, $n \longrightarrow \infty$
Если $\vec{\theta} = \{\theta_1, \ldots, \theta_r\}$, то $\hat{\vec{\theta}}(\vec{x_n}) \longrightarrow \vec{\theta}$; $(\hat{\theta}_i(\vec{x_n}) \longrightarrow \theta_i$, $i = 1, \ldots, r)$

Теорема (достаточные условия состоятельности оценки)
Если $M\hat{\theta}(\vec{x_n}) \longrightarrow \theta, n \longrightarrow \infty \wedge D\hat{\theta}(\vec{x_n}) \longrightarrow \theta, n \longrightarrow \infty$, то $\hat{\theta}(\vec{x_n})$ состоятельна.

Δ:

> $\alpha_n = M(\hat{\theta}(\vec{x_n}) - \theta$
>
> 1. $\forall h > 0 \exists n \geq n_0 \ |\alpha_n| = |M(\hat{\theta}(\vec{x_n})) - \theta| < h$
> 2. $|\hat{\theta}(\vec{x_n}) - \theta| = |\hat{\theta}(\vec{x_n}) - \theta - \alpha_n + \alpha_n| \leq |\hat{\theta}(\vec{x_n}) - \theta - \alpha_n| + |\alpha_n|$
> $P(|\hat{\theta}(\vec{x_n}) - \theta| < h) \geq P(|\hat{\theta}(\vec{x_n}) - \theta - \alpha_n| + |\alpha_n| < h)$
>   По неравенству Чебышева:
> $P(|\hat{\theta}(\vec{x_n}) - \theta| < h) \geq P(|\hat{\theta}(\vec{x_n}) - (\theta + \alpha_n)| < h + |\alpha_n|) \geq 1 - \frac{D(\hat{\theta}(\vec{x_n})}{h - |\alpha_n|)^2} \longrightarrow 1, \, n \longrightarrow \infty$
> $\implies P(|\hat{\theta}(\vec{x_n}) - \theta) \longrightarrow 1, \, n \longrightarrow \infty \implies \hat{\theta}(\vec{x_n}) \longrightarrow \theta, \, n \longrightarrow \infty$

□

2. Несмещенность

Опр. Оценка $\hat{\theta}(\vec{x_n})$ неизвестного параметра θ называется несмещенной, если $\forall n$ $M\hat{\theta}(\vec{x_n}) = \theta$, где $M\hat{\theta}(\vec{x_n})$ считается по всем возможным выборкам объема n.

Ассимптотическая несмещенность: $M(\theta(\hat{\vec{x_n}})) \longrightarrow 0, \, n \longrightarrow \infty$

Теорема Если $\hat{\theta}_1(\vec{x_n})$ и $\hat{\theta}_2(\vec{x_n})$ - две несмещенные оценки одного параметра с минимальной дисперсией, то $P(\hat{\theta}_1(\vec{x_n}) = \hat{\theta}_2(\vec{x_n})) = 1$

Δ:

> Пусть $D(\hat{\theta}_1(\vec{x_n})) = D(\hat{\theta}_2(\vec{x_n})) = d \leq D(\hat{\theta}_3(\vec{x_n}))$
> Возьмем $\hat{\theta}_3(\vec{x_n}) = \frac{1}{2}(\hat{\theta}_1(\vec{x_n}) + \hat{\theta}_2(\vec{x_n}))$
> $M\hat{\theta}_3 = \frac{1}{2}(M\hat{\theta}_1 + M\hat{\theta}_2) = \theta^2 \implies \hat{\theta}_3$ - несмещенная
> $D\hat{\theta}_3 = D(\frac{1}{2}(\theta_1 \hat{+} \theta_2)) = \frac{1}{4}(D\hat{\theta}_1 + D\hat{\theta}_2 + 2r\sqrt{D\hat{\theta}_1 D\hat{\theta}_2}) = (1)$
> $r(\xi, \eta) = \frac{cov(\xi, \eta)}{\sqrt{D\xi D\eta}}$
> $(1) = \frac{1}{4}(2d + 2rd) = \frac{d}{2}(1 + r) \geq d$
> Т.к. $|r| \leq 1 \implies$ выполняется "=" при $r = 1$
> $r = r(\hat{\theta}_1, \hat{\theta}_2) = 1 \implies \hat{\theta}_2 = A\hat{\theta}_1 + B$
>
> $$\begin{cases} M\hat{\theta}_2 = AM\hat{\theta}_1 + B \\ D\hat{\theta}_2 = A^2 D\hat{\theta}_1 \end{cases} \iff \begin{cases} \theta = A\theta + B \\ d = A^2 d \end{cases} \implies \begin{cases} A = 1 \\ B = 0 \end{cases} \implies P(\hat{\theta}_1 = \hat{\theta}_2) = 1$$

□