AI 中的数学

第十五、十六讲

方聪，概率统计部分参考章复熹和张原老师课件

2024 年秋季

1. 随机序列的收敛性
2. 大数定律与中心极限定理
3. 统计基本概念

§4.1 随机序列的收敛性

- 定义 1.1. 若

$$\lim_{n\to\infty} P\left(|\xi_n - \eta| \geqslant \varepsilon\right) = 0, \quad \forall \varepsilon > 0,$$

则称 ξ_n 依概率收敛于 η, 记为 $\xi_n \xrightarrow{P} \eta$.

- 定义 1.2. 若

$$P\left(\lim_{n\to\infty} \xi_n = \eta\right) = 1$$

则称 ξ_n 几乎必然收敛于 η, 记为 $\xi_n \xrightarrow{\text{a.s.}} \eta$.

- 定义 1.3. 若

$$\lim_{n\to\infty} F_{\xi_n}(x) = F_\eta(x), \quad \forall x \in C\left(F_\eta\right),$$

则称 ξ_n 依分布收敛于 η, 记为 $\xi_n \xrightarrow{d} \eta$.

- 定理 1.1. 若 $\xi_n \xrightarrow{\text{a.s.}} \eta$, 则 $\xi_n \xrightarrow{P} \eta$.
- 例 1.1 表明 $\xi_n \xrightarrow{P} \eta$ 不能推出 $\xi_n \xrightarrow{\text{a.s.}} \eta$.
- 定理 1.2. 若 $\xi_n \xrightarrow{P} \eta$, 则 $\xi_n \xrightarrow{d} \eta$.
- 例 1.2 表明 $\xi_n \xrightarrow{d} \eta$ 不能推出 $\xi_n \xrightarrow{P} \eta$.

- 几乎必然收敛，指不收敛的点 ω 是空集 (概率为 0)
- 依概率收敛要求随机变量不同的概率越来越小
- 弱收敛只关注分布与 ω 无关

例（不做要求）：对于 $(0,1)$ 上均匀分布，考虑下列随机变量序列：对任何正整数 k 及 $j = 1, \cdots, 2^k$，令

$$X_{k1} = \begin{cases} 1, & 0 < \omega < \frac{1}{2^k}, \\ 0, & \text{其他}; \end{cases} \quad , \quad X_{kj} = \begin{cases} 1, & \frac{j-1}{2^k} < \omega < \frac{j}{2^k}, \\ 0, & \text{其他}; \end{cases} \quad (j > 1).$$

这些 $X_{kj} : k \geqslant 1, j = 1, \cdots, 2^k\}$ 可排成一个序列：
$X_{11}, X_{12}, X_{21}, X_{22}, X_{23}, X_{24}, \cdots$（按照字典排列法，将第一个足标从小到大排，若相同则按第二个足标从小到大排），将该序列记为 $\xi_1, \xi_2, \cdots$，其中 $\xi_n = X_{k_n j_n}$，则对任何 $\varepsilon \in (0,1)$ 有：

$$P(|\xi_n| \geqslant \varepsilon) = P(\xi_n = 1) = \frac{1}{2^k_n}$$

在 $n \to \infty$ 时 $k_n \to \infty$，故有 $\lim_{n\to\infty} P(|\xi_n| \geqslant \varepsilon) = 0$，这表明 $\xi_n \xrightarrow{P} 0$.
而对于任何 $\omega \in (0,1)$，$\lim_{n\to\infty} \xi_n(\omega)$ 不存在。$\xi_n \xrightarrow{a.s.} \eta$ 不成立。

例：设随机变量 $X \sim N(0,1)$，令

$$\xi_{2n-1} = X, \quad \xi_{2n} = -X \quad (n = 1, 2, \cdots).$$

易知所有的 ξ_n 有相同的分布函数 $\phi(x)$，为标准正态分布函数，显然 $\xi_n \xrightarrow{\omega} \eta$. 但是对 $\varepsilon > 0$，有

$$P(|\xi_n - X| \geqslant \varepsilon) = \begin{cases} 0, & n\text{是奇数}, \\ P(|X| \geqslant \frac{\varepsilon}{2}), & n\text{是偶数}. \end{cases}$$

可见 $\xi_1, \xi_2, \cdots$ 并不依概率收敛于 X。

- 假设 $X_1, X_2, \cdots$ 是随机变量序列，令

$$S_n = X_1 + \cdots + X_n.$$

- 定义 1.4. 若 $EX_n, n = 1, 2, \cdots$ 都存在，且

$$\frac{1}{n}(S_n - ES_n) \xrightarrow{P} 0$$

  则称 $X_1, X_2, \cdots$ 服从 (弱) 大数律 (Weak Law of Large Numbers, WLLN)

- 定义 1.5. 若 $EX_n, n = 1, 2, \cdots$ 都存在，且

$$\frac{1}{n}(S_n - ES_n) \xrightarrow{\text{a.s.}} 0,$$

  则称 $X_1, X_2, \cdots$ 服从强大数律 (SLLN).

定义：若对任意 $n \geqslant 2$ 都有 $X_1, \cdots, X_n$ 相互独立，则称 $X_1, X_2, \cdots$ 是相互独立的随机变量序列。

若 $X_1, X_2, \cdots$ 相互独立，且 $X_n \stackrel{d}{=} X_1, \forall n \geqslant 2$，则称 $X_1, X_2, \cdots$ 独立同分布，记为 i.i.d. (independent and identically distributed).

例：设随机变量序列 $\{X_n\}$ 独立同分布，其密度函数为

$$p(x)=\begin{cases}\frac{1}{\beta}, & 0<x<\beta,\\ 0, & \text{其他}.\end{cases}$$

其中常数 $\beta>0$，令 $Y_n=\max\{X_1,X_2,\cdots,X_n\}$，证明 $Y_n\xrightarrow{P}\beta$.

例：设随机变量序列 $\{X_n\}$ 独立同分布，其密度函数为

$$p(x)=\begin{cases}\frac{1}{\beta}, & 0<x<\beta,\\ 0, & \text{其他}.\end{cases}$$

其中常数 $\beta>0$，令 $Y_n=\max\{X_1,X_2,\cdots,X_n\}$，证明 $Y_n\xrightarrow{P}\beta$.

证明：因为当 $x<0$ 时，有 $P(Y_n\leqslant x)=0$，当 $x\geqslant\beta$ 时，有 $P(Y_n\leqslant x)=1$，当 $0\leqslant x<\beta$ 时，有

$$P(Y_n\leqslant x)=\prod_{i=1}^{n}P(X_i\leqslant x)=\prod_{i=1}^{n}\int_0^x\frac{1}{\beta}dx=\left(\frac{x}{\beta}\right)^n,$$

所以对任意的 $\varepsilon>0(\varepsilon<\beta)$，当 $n\to\infty$ 时，有

$$P(|Y_n-\beta|\geqslant\varepsilon)=P(Y_n\leqslant\beta-\varepsilon)=\left(\frac{\beta-\varepsilon}{\beta}\right)^n\to 0,$$

所以有 $Y_n\xrightarrow{P}\beta$.

定理

Chebyshev's WLLN, 定理 *2.1* 假设 $X_1, X_2, \cdots$ 相互独立, 且 $\mathrm{var}(X_i) \leqslant M, \forall i$. 那么,

$$\frac{1}{n}(S_n - ES_n) \xrightarrow{P} 0.$$

- 令 $A_n = \left\{\left|\frac{1}{n}(S_n - ES_n)\right| \geqslant \varepsilon\right\}$. 需验证 $P(A_n) \to 0$.
- 由切比雪夫不等式,

$$\begin{aligned} P(A_n) &= P(|S_n - ES_n| \geqslant n\varepsilon) \leqslant \frac{1}{(n\varepsilon)^2}\mathrm{var}(S_n) \\ &\leqslant \frac{nM}{n^2\varepsilon^2} = \frac{M}{\varepsilon^2} \cdot \frac{1}{n} \to 0. \end{aligned}$$

- “相互独立”可减弱为“两两不相关”.

推论：设 $X_1, X_2, \cdots$ 独立同分布，$\mathrm{var}(X_1) < \infty$，则当 $n \to \infty$ 时，

$$\frac{S_n}{n} \xrightarrow{P} E(X_1).$$

推论：（伯努利大数律）单次试验中 A 发生的概率为 p，设在 n 次试验中事件 A 发生了 ν_n 次，则当 $n \to \infty$ 时，

$$\frac{\nu_n}{n} \xrightarrow{P} p.$$

推论：设 $X_1, X_2, \cdots$ 独立同分布, $\mathrm{var}(X_1) < \infty$, 则当 $n \to \infty$ 时，

$$\frac{S_n}{n} \xrightarrow{P} E(X_1).$$

推论：（伯努利大数律）单次试验中 A 发生的概率为 p, 设在 n 次试验中事件 A 发生了 ν_n 次，则当 $n \to \infty$ 时，

$$\frac{\nu_n}{n} \xrightarrow{P} p.$$

证明：令

$$X_i = \begin{cases} 1, & \text{第}i\text{次试验中}A\text{发生}, \\ 0, & \text{第}i\text{次试验中}A\text{不发生}, \end{cases} \quad (i = 1, 2, \cdots),$$

则 $\frac{\nu_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 由于 $X_1, X_2, \cdots$ 是独立同分布的随机变量序列， $E(X_i) = p$， $\mathrm{var}(X_i) = p(1-p)(i = 1, 2, \cdots)$，故由上一推论知本推论成立。

如果不假定 $E(X_i)$ 存在，上述推论是否成立？

例：设 $X_1, X_2, \cdots$ 独立同分布，密度为 $p(x) = \frac{1}{\pi(x^2+1)}$. 可以证明，$\frac{S_n}{n}$ 与 X_1 有相同的密度. 于是，对任何 a 和 $\varepsilon > 0$，有

$$P\left(\left|\frac{S_n}{n} - a\right| > \varepsilon\right) = P\left(|X_1 - a| > \varepsilon\right) \text{ 不趋于0.}$$

故 $\frac{S_n}{n}$ 不能以概率收敛于 a。

定理

Cantelli's SLLN, 定理 *2.2*, 引理 *2.1* 假设 $X_1, X_2, \cdots$ 相互独立, EX_i 存在, 且 $E(X_i - EX_i)^4 \leqslant M, \forall i$. 那么,

$$\frac{1}{n}(S_n - ES_n) \xrightarrow{a.s.} 0.$$

- 推论 2.3. 设 $X_1, X_2, \cdots$ 独立同分布, EX_1^4 存在, 则 $\frac{S_n}{n} \xrightarrow{\text{a.s.}} EX_1$.
- 推论 2.4. 单次小试验中事件 A 发生的概率为 p. 在独立重复试验中, 前 n 次试验中 A 发生的频率 $\xrightarrow{\text{a.s.}} p$.

- 定理 2.4. (Kolmogorov's SLLN). 假设 $X_1, X_2, \cdots$ 独立同分布, 期望存在, 则 $\frac{1}{n} S_n \xrightarrow{\text{a.s.}} EX_1$.
- 时间平均 = 空间平均, (期望的含义).

例：设 $X_1, X_2, \cdots$ 是相互独立的随机变量序列，$X_1 \equiv 0$，对一切 $n \geqslant 2$，X_n 只取三个可能值 $n, -n, 0$，且

$$P(X_n) = n = P(X_n = -n) = \frac{1}{2n \ln n}, \quad P(X_n = 0) = 1 - \frac{1}{n \ln n}.$$

证明 $X_1, X_2, \cdots$ 服从切比雪夫大数定律。

例：设 $X_1, X_2, \cdots$ 是相互独立的随机变量序列，$X_1 \equiv 0$，对一切 $n \geqslant 2$，X_n 只取三个可能值 $n, -n, 0$，且

$$P(X_n) = n = P(X_n = -n) = \frac{1}{2n\ln n}, \quad P(X_n = 0) = 1 - \frac{1}{n\ln n}.$$

证明 $X_1, X_2, \cdots$ 服从切比雪夫大数定律。

证明：

$$E(X_n) = 0, \quad \operatorname{var}(X_1) = 0, \quad \operatorname{var}(X_n) = \frac{n}{\ln n}(n = 2, 3, \cdots).$$

令 $S_n = \sum_{i=1}^n X_i$，则 $\operatorname{var}(S_n) = \sum_{i=1}^n \operatorname{var}(X_i) = \sum_{i=2}^n \frac{i}{\ln i}$. 由于 $x \geqslant 3$ 时 $\frac{x}{\ln x}$ 是 x 的增函数，故 $\operatorname{var}(S_n) \leqslant \frac{2}{\ln 2} + \frac{n^2}{\ln n}$，利用切比雪夫不等式，有

$$P\left(\left|\frac{S_n - E(S_n)}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{1}{n^2\varepsilon}\operatorname{var}(S_n) \to 0 \quad (\varepsilon > 0, n \to \infty).$$

这表明 $X_1, X_2, \cdots$ 服从切比雪夫大数定律。

应用 (1): 统计方法的理论依据.

- 数据: $X_1, \cdots, X_n$ 为 X 的 n 次独立观测值, 它们独立同分布.
- 估计期望:

$$\bar{X} := \frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow{\text{a.s.}} EX.$$

- 估计方差:

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - (\bar{X})^2 \xrightarrow{\text{a.s.}} EX^2 - (EX)^2 = \mathrm{var}(X).$$

应用 (2): 计算机模拟期望、概率.

- 例 2.3. 设有 m 枚炮弹同时射击, 第 i 枚炮弹落点为 (x_i, y_i),

$$\varphi\left(x_1, y_1; \cdots; x_m, y_m\right) = \begin{cases} 1, & \text{若落点造成有效毁伤;} \\ 0, & \text{否则.} \end{cases}$$

- 设第 i 枚炮弹的瞄准点为 (a_i, b_i), 实际落点 (X_i, Y_i). 模型假设: $X_1, \cdots, X_n; Y_1, \cdots, Y_n$ 相互独立, 且

$$X_i \sim N\left(a_i, \sigma_1^2\right), \quad Y_i \sim N\left(b_i, \sigma_2^2\right).$$

- SLLN:

$$\begin{aligned} & P\left(\varphi\left(X_1, Y_1; \cdots; X_m, Y_m\right) = 1\right) \\ \approx & \frac{1}{n} \sum_{k=1}^{n} \varphi\left(X_1^{(k)}, Y_1^{(k)}; \cdots; X_m^{(k)}, Y_m^{(k)}\right). \end{aligned}$$

应用 (3)：估计积分 $I = \int_a^b f(x)dx$.

- $I = \int_0^1 f(a+(b-a)u)(b-a)du$, 因此不妨假设 $a=0, b=1$.
- $I = \int_0^1 f(x) \cdot 1dx = Ef(U)$.
- SLLN:

$$I \approx \frac{1}{n}\sum_{i=1}^{n}\left(f\left(U_1\right)+\cdots+f\left(U_n\right)\right)$$

- 定义 1.6. 若 $EX_n, \mathrm{var}(X_n), n = 1, 2, \cdots$ 都存在, $\mathrm{var}(X_n)$ 不全为 0，且

$$S_n^* = \frac{S_n - ES_n}{\sqrt{\mathrm{var}(S_n)}} \xrightarrow{d} Z \sim N(0, 1),$$

则称 $X_1, X_2, \cdots$ 服从中心极限定理 (Central Limit Theorem, CLT).

- 定理 3.1(Linderberg-Levy CLT). 假设 $X_1, X_2, \cdots$ 独立同分布, $0 < \operatorname{var}(X_1) < \infty$. 那么,

$$S_n^* \xrightarrow{d} Z \sim N(0,1).$$

- 应用:

$$P(S_n \leqslant x) = P(S_n^* \leqslant x^*) \approx \Phi(x^*) =: p,$$

其中, $x^* = \frac{x - n\mu}{\sqrt{n}\sigma}$.

例 3.1. 加法器同时收到 20 个噪声电压 $V_k, k=1,\cdots,20$, 它们独立同分布, $V_1 \sim U(0,10)$. 记 $V=\sum_{k=1}^{20} V_k$, 求 $P(V>105)$.

- 此题已知 $n=20, x=105$, 求 p.
- $EV_1=5, \operatorname{var}(V_1)=\frac{10^2}{12}$.
- 根据 CLT,

$$P(V>105)=P\left(V^*>x^*\right) \approx 1-\Phi\left(x^*\right)=: p,$$

其中,

$$x^*=\frac{105-20*5}{\sqrt{20*\frac{100}{12}}}.$$

- 计算得 $x^* \approx 0.387$. 查表得 $\Phi\left(x^*\right)=0.652$, 从而所求之 $p=1-0.652=0.348$.

例 3.2. 旅馆有 500 间客房, 每间有一台 2 千瓦的空调. 入住率为 80%. 问: 需多少千瓦的电力能有 99% 的把握保证电力足够?

例 3.2. 旅馆有 500 间客房, 每间有一台 2 千瓦的空调. 入住率为 80%. 问: 需多少千瓦的电力能有 99% 的把握保证电力足够?

- 已知 n, p, 求 x. 假设提供 x 千瓦.
- $A_i =$ 第 i 间房开空调, $P(A_i) = 80\%, X_i = 2 \times 1_{A_i}, n = 500$.
- $EX_1 = 2 \times 0.8 = 1.6, \operatorname{var}(X_1) = 4 \times 0.8 \times 0.2 = 0.64$.
- 要求 x 满足: $P(S_n \leqslant x) \geqslant 99\% = p$. 根据 CLT, 要求

$$P(S_n^* \leqslant x^*) \approx \Phi(x^*) \geqslant 0.99$$

其中

$$x^* = \frac{x - 500 * 2 * 0.8}{\sqrt{500 * 2^2 * 0.8 * 0.2}}.$$

- 查表得 $\Phi(2.33) = 0.99$. 即, 要求 $x^* \geqslant 2.33$.
即, 要求 $x \geqslant 800 + 2.33 * \sqrt{320} = 841.68$, 从而需 842 千瓦.

例 3.3. 桥的强度 $Y \sim N(300, 40)$ (单位: 吨), 车的平均重量为 5, 方差为 2 . 问: 为保证桥不出问题的概率不小于 0.99997, 最多允许在桥上同时出现多少辆车?

例 3.3. 桥的强度 $Y \sim N(300, 40)$ (单位: 吨), 车的平均重量为 5, 方差为 2 . 问: 为保证桥不出问题的概率不小于 0.99997, 最多允许在桥上同时出现多少辆车?

- 已知 $x, p = 0.99997$, 求 n.
- 假设有 n 辆车在桥上, 第 i 辆的重量为 X_i.
  模型假设 $X_1, \cdots, X_n$ 独立同分布, 且与 Y 相互独立.
- 记 $S_n = X_1 + \cdots + X_n$, 保证 $P(S_n \leqslant Y) \geqslant 0.99997$.
- 根据 CLT 及独立性, 近似地, $S_n - Y \sim N(5n - 300, 2n + 40)$.
  因此, 令 $0^* = \frac{0-(5n-300)}{\sqrt{2n+40}}$, 有

$$P(S_n \leqslant Y) = P(S_n - Y \leqslant 0) \approx \Phi(0^*).$$

- 要求 $\Phi(0^*) \geqslant 0.9997$, 查表得 $x^* \geqslant 4$, 即, $\frac{-(5n-300)}{\sqrt{2n+40}} \geqslant 4$, 解得 $n \leqslant 50.5$. 从而, 最多同时 50 辆车.

- 参考习题书：茆诗松，程依明，濮晓龙. 概率论与数理统计教程. 高等教育出版社.
- 作业都会讲
- 考试难度与作业相当，最后依情况给出考点范围与复习题
- 应用：统计是概率的应用，机器学习是统计的应用

- 数理统计学：收集、整理、分析带随机性的数据 (data).
- 例如，学校希望了解两名学生的学习能力有没有差别，他们的成绩表如下.

甲	80	70	60	85
乙	70	90	65	80

统计学（数据科学）：从数据中发现规律！

定义：所考察的对象的总和称为总体，在统计学中可以归结为随机变量或其他形式的随机量。

例如，考察电子产品的使用寿命，于是将所有电子产品的使用寿命作为总体。所谓总体特性，就是使用寿命的特性，或者是刻画使用寿命的随机变量 X 的特性，该随机变量的分布称为总体分布。可以假定 X 的分布为指数分布，其分布密度有下列形式：

$$p(x,\theta) = \frac{1}{\theta}\exp\left\{-\frac{x}{\theta}\right\} \quad (x>0,\theta>0),$$

式中 θ 是分布的参数。

- 设用 $F(x,\theta)$ 表示随机变量 X 分布密度相应的分布函数，用 F_θ 表示相应的分布
- 为获取分布 F_θ 的信息，假定 F_θ 属于一个分布族，用 $\mathcal{F}=\{F_\theta,\theta\in\Theta\}$ 表示这个分布族
- θ 称为参数，Θ 称为参数空间
- $X\sim F_\theta\in\mathcal{F}$ 形成了这个统计问题的模型，称为总体模型

例如，电子产品的使用寿命 X 的分布 F_θ 由分布密度确定，其中参数 $\theta\in(0,+\infty)$。当 θ 确定后，我们获得了电子产品使用寿命的全部信息。

- 总体模型只涉及 X 这个随机变量，而没有涉及数据
- 观察数据 $\mathbf{x} = (x_1, \cdots, x_n)$ 是 $\mathbf{X} = (X_1, \cdots, X_n)$ 的观察值，独立同分布 F_θ 产生
- $\mathbf{X} = (X_1, \cdots, X_n)$ 为样本，称 n 为样本量，称 $\mathbf{X}$ 的取值 $\mathbf{x} = (x_1, \cdots, x_n)$ 为样本值

称 $\mathbf{X}$ 的所有可能取值的集合为样本空间 $\mathcal{X}$，在样本空间上的分布为 P_θ，称 $\mathbf{X} \sim P_\theta(\theta \in \Theta)$ 为统计模型。

- 总体模型只涉及 X 这个随机变量，而没有涉及数据
- 观察数据 $\mathbf{x} = (x_1, \cdots, x_n)$ 是 $\mathbf{X} = (X_1, \cdots, X_n)$ 的观察值，独立同分布 F_θ 产生
- $\mathbf{X} = (X_1, \cdots, X_n)$ 为样本，称 n 为样本量，称 $\mathbf{X}$ 的取值 $\mathbf{x} = (x_1, \cdots, x_n)$ 为样本值

称 $\mathbf{X}$ 的所有可能取值的集合为样本空间 $\mathcal{X}$，在样本空间上的分布为 P_θ，称 $\mathbf{X} \sim P_\theta(\theta \in \Theta)$ 为统计模型。

$\mathbf{X}$ 着重方法/理论分析; $\mathbf{x}$ 着重应用/计算.

- 模型的参数 θ 可以是常数向量或者其他的量
- 一旦参数的值确定后，统计模型中的分布就完全确定了
- 在某些统计问题中，需要了解与参数有关的量，即 θ 的函数 $g(\theta)$，为了简便，将 $g(\theta)$ 也称为参数

例：（测量问题）对某待估量 a 重复独立测量 n 次，得到测量值 $x_1, \cdots, x_n$.

- 测量值带有误差，总体分布 $X = a + e$，其中 $e \sim N\left(0, \sigma^2\right)$. 即，$X \sim N\left(a, \sigma^2\right)$
- 相应的参数空间为 $\Theta = \{\theta = \left(a, \sigma^2\right) : a \in \mathbb{R}, \sigma^2 > 0\}$
- 参数 $\theta = \left(a, \sigma^2\right)$. 其中，$\sigma^2$ 不是所关心的，称为讨厌参数
- $P_\theta : \vec{X}$ 的联合密度为

$$f_\theta(x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_i - a)^2}{2\sigma^2}\right\}.$$

- $\mathbf{X} \sim P_\theta$，$\theta = (a, \sigma^2) \in \Theta$ 形成了统计模型
- 研究对象 θ 或 $g(\theta)$. 例，$g\left(a, \sigma^2\right) = a$

定义：设 $\mathbf{X} \sim P_\theta(\theta \in \Theta)$ 是一个统计模型，则定义在样本空间 $\mathcal{X}$ 上的任何函数 $T(\mathbf{x})(\mathbf{x} \in \mathcal{X})$ 都称为统计量。

- 统计量是一个只依赖数据的函数，当 $(x_1, \cdots, x_n)$ 的值给定后根据函数关系可以算出 $T(\mathbf{x})$ 的值
- $T(\mathbf{X})$ 还是一个随机变量，具有分布，且在不同参数值下具有不同的分布。严格意义下，统计量具有分布族

定义：设 $\mathbf{X} \sim P_\theta(\theta \in \Theta)$ 是一个统计模型，则定义在样本空间 $\mathcal{X}$ 上的任何函数 $T(\mathbf{x})(\mathbf{x} \in \mathcal{X})$ 都称为统计量。

- 统计量是一个只依赖数据的函数，当 $(x_1, \cdots, x_n)$ 的值给定后根据函数关系可以算出 $T(\mathbf{x})$ 的值
- $T(\mathbf{X})$ 还是一个随机变量，具有分布，且在不同参数值下具有不同的分布。严格意义下，统计量具有分布族

例如在测量问题中，最常见的统计量为样本均值
$T = \frac{1}{n}(X_1 + \cdots + X_n)$，当观察值为 $x_1, \cdots, x_n$ 时，
$T = \frac{1}{n}(x_1 + \cdots + x_n)$ 为一个数值

- 当 $\mathbf{X} = (X_1, \cdots, X_n)$ 时，统计量是样本的函数，为随机变量
- T 的分布，$T \sim N\left(a, \frac{\sigma^2}{n}\right)$
- 注意统计量的分布含有未知参数 (a, σ^2)

基本问题

估计问题：依赖于样本的统计量就可以作为参数 a 的估计，在估计问题中，估计参数的统计量也称为估计量。

基本问题

估计问题：依赖于样本的统计量就可以作为参数 a 的估计，在估计问题中，估计参数的统计量也称为估计量。

例：（测量问题续）测量问题中待测量 a 的一个估计为 $T_1=\frac{1}{n}\sum_{i=1}^{n}x_i$。当 $(X_1,\cdots,X_n)$ 服从多元正态分布时，其常系数线性组合的分布也是正态分布，利用 $X_i(i=1,\cdots,n)$ 独立同分布的特性，计算 T_1 的期望和方差，可得

$$E(T_1)=\frac{1}{n}\left[\sum_{i=1}^{n}E(X_i)\right]=a,$$

$$\mathrm{var}(T_1)=\frac{1}{n^2}\left[\sum_{i=1}^{n}\mathrm{var}(X_i)\right]=\frac{\sigma^2}{n},$$

这样我们得到 $T_1\sim N\left(a,\frac{\sigma^2}{n}\right)$。

假设检验：对假设 H_0 回答“是”或“否”。

例如，规定不合格率不能超过 3%. 现有 200 件产品，从中任意抽取 10 件，发现 2 件不合格. 问：是否可以出厂？

线性回归：研究变量 Y 对 x 的线性依赖关系，

$$Y = b_0 + b_1 x + e, \quad e \sim N\left(0, \sigma^2\right).$$