

定义 (随机变量)

设 $\xi(\omega)$ 是定义在概率空间 (Ω, \mathcal{F}, P) 上的单值实函数, 且对于任一 Borel 集 B , 有

原像属于事件域

$$\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathcal{F}, \quad (2.1.1)$$

就称 $\xi(\omega)$ 为随机变量 (random variable), 称 $P(\{\omega : \xi(\omega) \in B\})$, $B \in \mathcal{B}$ (一维 Borel σ -域), 为随机变量 $\xi(\omega)$ 的概率分布 (probability distribution).

2.1.2 离散型随机变量

定义 (离散型随机变量)

若随机变量 ξ 可能取的实数值至多可列个 (有限个或可列个), 则称 ξ 为离散型随机变量 (discrete random variables).

对于离散型随机变量 ξ , 需要关心两方面的内容:

(1) 可能的取值: $x_i, i = 1, 2, \dots, n (n \leq \infty)$;

(2) ξ 取这些值的概率大小:

$$p_i := p(x_i) := P(\xi = x_i), \quad i = 1, 2, \dots, n.$$

称

$$P(\xi = x_i) = p(x_i), \quad i = 1, 2, \dots, n$$

为 ξ 的分布列 (distribution sequence) 或分布律 (distribution law), 有时也称为 ξ 的概率分布.

分布列的性质 ($n \leq \infty$):

(1) 非负性: $p(x_i) \geq 0, \quad i = 1, \dots, n;$

(2) 规范性: $\sum_{i=1}^n p(x_i) = 1.$

注 有了分布列, 就可以确定离散型随机变量 ξ 的整个概率分布: 由概率的有限/可列可加性, 对任一 Borel 集 B , 有

$$\begin{aligned} P(\xi \in B) &= P(\{\omega : \xi(\omega) \in B\}) \\ &= P\left(\bigcup_{x_i \in B} \{\omega : \xi(\omega) = x_i\}\right) \\ &= \sum_{x_i \in B} p(x_i). \end{aligned}$$

一些常见的、重要的离散型随机变量:

1. 退化分布 (degenerate distribution)

定义

若随机变量 ξ 只取一个实数值 c , 即分布列为

$$P(\xi = c) = 1,$$

则称它为退化分布, 又称为单点分布或 *Dirac* 分布.

2. 两点分布 (two-point distribution)

定义

若随机变量 ξ 只取两个实数值: x_1, x_2 , 且相应的分布列为

ξ	x_1	x_2
P	p	q

其中 $p, q > 0, q = 1 - p$, 则称 ξ 服从两点分布.

3. 二项分布 (binomial distribution)

定义

若随机变量 ξ 的分布列为

$$P(\xi = k) = C_n^k p^k q^{n-k} =: b(k; n, p), \quad k = 0, 1, \dots, n,$$

其中 $p, q > 0, p + q = 1$, 则称 ξ 服从参数为 (n, p) 的二项分布, 记作 $\xi \sim B(n, p)$.

注 (1) $n = 1$ 时的二项分布即为 $0-1(p)$ 分布. (2) 二项分布随机变量通常是从 n 重 Bernoulli 试验中产生的.

二项分布的性质:

(1) $b(k; n, p) = b(n - k; n, 1 - p)$. 这是因为 $C_n^k = C_n^{n-k}$, 或理解成: 在 n 重 Bernoulli 试验中, $\{k \text{ 次成功}\}$ 的概率 = $\{n - k \text{ 次失败}\}$ 的概率.

(2) 单调增减性以及最可能成功次数.

固定 n, p . 由于

$$\frac{b(k; n, p)}{b(k-1; n, p)} = \frac{C_n^k p^k (1-p)^{n-k}}{C_n^{k-1} p^{k-1} (1-p)^{n-k+1}} = 1 + \frac{(n+1)p - k}{k(1-p)},$$

因此

当 $k < (n+1)p$ 时, $\frac{b(k; n, p)}{b(k-1; n, p)} > 1$, $b(k; n, p)$ 单调增加;

当 $k > (n+1)p$ 时, $\frac{b(k; n, p)}{b(k-1; n, p)} < 1$, $b(k; n, p)$ 单调减少.

- 当 $(n+1)p$ 是整数时,

$$b(k; n, p) = b(k-1; n, p), \quad k = (n+1)p.$$

$(n+1)p$ 和 $(n+1)p-1$ 为最可能成功次数.

- 当 $(n+1)p$ 不是整数时, 由上面的单调性分析以及

$$b(\lfloor (n+1)p \rfloor + 1; n, p) < b(\lfloor (n+1)p \rfloor; n, p)$$

可知 $\lfloor (n+1)p \rfloor$ 为最可能成功次数 (这里, $\lfloor \cdot \rfloor$ 表示取整符号).

(3) 递推公式.

设 $\xi \sim B(n, p)$, 则

$$P(\xi = k + 1) = \frac{p(n - k)}{q(k + 1)} P(\xi = k).$$

此公式容易由二项分布的表达式得到. 从 $P(\xi = 0) = q^n$ 出发, 可以用此公式递推得到 $\{P(\xi = k), k = 0, \dots, n\}$ 的值.

也可以使用各种数学和统计软件, 如 Matlab, SPSS, R, Python 等, 计算概率分布值.

(4) $n \rightarrow \infty$ 时的渐近性质.

假定 p 与 n 有关, 记作 p_n . 有下列的“二项分布的泊松 (Poisson) 定理”.

定理 (Poisson 定理)

如果存在正常数 λ , 当 $n \rightarrow \infty$ 时, 有 $np_n \rightarrow \lambda$, 则

$$\lim_{n \rightarrow \infty} b(k; n, p_n) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

证明 利用初等不等式

$$|a^n - b^n| \leq n|a - b|, \quad \text{如果 } |a| \leq 1, |b| \leq 1,$$

可知当 $n \rightarrow \infty$ 时,

$$\left| (1 - p_n)^n - \left(1 - \frac{\lambda}{n}\right)^n \right| \leq n \left| p_n - \frac{\lambda}{n} \right| \rightarrow 0.$$

记 $\lambda_n = np_n$, 则 $p_n = \lambda_n/n$. 固定 k , 可得

$$\begin{aligned} b(k; n, p_n) &= C_n^k p_n^k (1 - p_n)^{n-k} \\ &= \frac{n(n-1) \cdots (n-k-1)}{k!} \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\ &= \frac{\lambda_n^k}{k!} \cdot \frac{n(n-1) \cdots (n-k-1)}{n^k} \cdot \left(1 - \frac{\lambda_n}{n}\right)^n \cdot \left(1 - \frac{\lambda_n}{n}\right)^{-k} \\ &\rightarrow \frac{\lambda^k}{k!} e^{-\lambda} \quad (n \rightarrow \infty). \end{aligned}$$

注 通常, p 与 n 无关. 实践表明: 当 n 很大 ($n \geq 50$), p 很小 ($p \leq 0.1$), 而 np 不很大时, 可取 $\lambda = np$, 且

$$b(k; n, p) \approx \frac{\lambda^k}{k!} e^{-\lambda}.$$

$\frac{\lambda^k}{k!} e^{-\lambda}$ 的计算要比 $b(k; n, p)$ 的计算容易得多.

Poisson 分布的应用:

- (1) 二项分布的近似计算 (见 Poisson 定理).
- (2) 可以用来描述离散型随机现象: 如果 n 个独立事件 A_1, A_2, \dots, A_n 中每个发生的概率 p 很小, 那么这 n 个事件发生的次数近似服从 Poisson 分布 $P(np)$. (虽然这里不一定是 n 次的重复试验, 但这 n 个事件发生的次数仍服从 $B(n, p)$)
- (3) 通常认为单位时间/区间/面积/体积/ \dots 中的计数过程服从 Poisson 分布. 例如, 一定时间内接到的电话呼叫数、某公交站台在一定时间内的上车人数、一本书的错别字个数、一平方米玻璃上的气泡数等等, 均可认为服从 Poisson 分布.

假设 ξ 为某放射性物质在时间区间 $(0, t]$ 内放射出的 α 粒子数. 将时间区间 $(0, t]$ 等分成 n 个小区间, 每个小区间的长度为 t/n . 假设:

- (1) 每个小区间内最多只有一个 α 粒子放射出来, 并且放射出一个 α 粒子的概率为 $p_n = \lambda t/n$, 这里的 λ 是某正常数, 用来刻画放射强度;
- (2) 各个小区间内是否放射出 α 粒子是相互独立的.

在上述假定下, 这个放射性物质在时间区间 $(0, t]$ 内放射出来的 α 粒子数 ξ 的概率分布是 $B(n, p_n)$ 的极限分布. 由于 $\lim_{n \rightarrow \infty} np_n = \lambda t$, 根据 Poisson 定理, $\xi \sim P(\lambda t)$.

这就是把单位时间/区间/面积/体积/... 中的计数过程当成 Poisson 分布来处理的理由.

具体实例 1 1910 年科学家卢瑟福 (Rutherford) 和盖革 (Geiger) 观测了放射性物质钋 (Polonium) 放射 α 粒子的情况: 他们进行了 $N = 2608$ 次观测, 每次观测 7.5 秒, 一共观测到 10094 个 α 粒子, 下表是实验数据的拟合情况 (参数 $\lambda = 3.87$):

放射粒子数 ξ	观测到的次数 n_k	频率 $f_k = \frac{n_k}{N}$	概率 $P(\xi = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
0	57	0.022	0.021
1	203	0.078	0.081
2	383	0.147	0.156
3	525	0.201	0.201
4	532	0.204	0.195
5	408	0.156	0.151
6	273	0.105	0.097
7	139	0.053	0.054
8	45	0.017	0.026
9	27	0.010	0.011
10	10	0.004	0.004
≥ 11	6	0.002	0.003
总计	$N = 2608$	0.999	1.000

具体实例 2 有人统计了 1500–1931 年共 $N = 432$ 年间的战争, 一年中爆发战争的频率与 Poisson 分布的概率接近, 下表是拟合情况 (参数 $\lambda = 0.69$):

年度战争次数 ξ	观测到的年数 n_k	频率 $f_k = \frac{n_k}{N}$	概率 $P(\xi = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
0	223	0.516	0.502
1	142	0.329	0.346
2	48	0.111	0.119
3	15	0.035	0.028
≥ 4	4	0.009	0.005
总计	$N = 432$	1.000	1.000

进一步放宽条件: 一般地, 设有 n 个事件 A_1, A_2, \dots, A_n , 第 i 个事件发生的概率为 p_i . 如果这些 p_i 都很小, 而且这些事件相互独立或者近似独立, 那么这些事件发生的次数近似服从 Poisson 分布 $P(\lambda)$, 其中

$$\lambda = \sum_{i=1}^n p_i.$$

5. 几何分布 (geometric distribution)

定义

若随机变量 ξ 的分布列为

$$P(\xi = k) = q^{k-1}p, \quad k = 1, 2, \dots,$$

其中 $p, q > 0, p + q = 1$, 则称 ξ 服从参数为 p 的几何分布, 记作 $\xi \sim \text{Geo}(p)$.

注 (1) 在 Bernoulli 概型中, 若单次试验成功的概率为 p , 则直到首次成功的试验次数服从参数为 p 的几何分布. (2) $r = 1$ 时的 Pascal 分布为几何分布.

定理

若 ξ 是取正整数的随机变量, 且具有无记忆性, 则 ξ 服从几何分布.

*证 假设 ξ 具有无记忆性. 记

$$p = P(\xi = k + 1 | \xi > k), \quad q_k = P(\xi > k), \quad p_k = P(\xi = k).$$

那么 $p_{k+1} = q_k - q_{k+1}$, 而且在已知 $\xi > k$ 的条件下 $\xi = k + 1$ 的条件概率为 p_{k+1}/q_k . 因此

$$\frac{p_{k+1}}{q_k} = p.$$

即

$$\frac{q_{k+1}}{q_k} = 1 - p.$$

注意到 $q_0 = 1$, 那么 $q_k = (1 - p)^k$. 因此,

$$p_k = (1 - p)^{k-1}p, \quad k = 1, 2, \dots,$$

这正是几何分布的分布列.

6. 超几何分布 (hyper-geometric distribution)

定义

若随机变量 ξ 的分布列为

$$P(\xi = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \quad k = \max\{0, n - (N - M)\}, \dots, \min(n, M),$$
$$(n \leq N, M \leq N)$$

则称 ξ 服从参数为 (n, M, N) 的超几何分布, 记作 $\xi \sim H(n, M, N)$.

注 超几何分布的一个典型例子: 在产品质量的不放回抽样中, 若 N 件产品中有 M 件次品, 则抽检 n 件时所得次品数服从超几何分布.

超几何分布的性质:

若 n, k 不变, $N \rightarrow \infty, M/N \rightarrow p$, 则超几何分布可用二项分布来逼近.
即,

$$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \rightarrow C_n^k p^k q^{n-k} \quad (N \rightarrow \infty).$$

(产品充分多时, 有放回抽取和不放回抽取没有本质差别)

事实上,

$$\begin{aligned} & \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \\ &= \frac{M!}{k!(M-k)!} \cdot \frac{(N-M)!}{(n-k)!(N-M-n+k)!} \cdot \frac{n!(N-n)!}{N!} \\ &= C_n^k \cdot \frac{(M-k+1) \cdots M}{(N-k+1) \cdots N} \cdot \frac{(N-M-n+k+1) \cdots (N-M)}{(N-n+1) \cdots (N-k)} \\ &\rightarrow C_n^k p^k q^{n-k} \quad (N \rightarrow \infty). \end{aligned}$$

定义 (分布函数)

设 ξ 为概率空间 (Ω, \mathcal{F}, P) 上的随机变量. 称

$$F(x) = P(\xi \leq x), \quad x \in \mathbb{R}$$

为随机变量 ξ 的分布函数.