

# Probability

Wenxiao Yang\*

\*Department of Mathematics, University of Illinois at Urbana-Champaign

2021

## 目录

<b>1</b>	<b>Distribution</b>	<b>2</b>
1.1	Bernoulli Distribution Bernoulli( $\pi$ ): 事件发生概率为 $\pi$ . . . . .	2
1.2	Binomial distribution $bin(n, \pi)$ : $n$ 次 Bernoulli . . . . .	2
1.3	Multinomial Distribution . . . . .	2
1.4	Poisson Distribution $Pois(\lambda)$ : 单位时间发生 $k$ 次事件的概率 . . . . .	2
1.4.1	Sums of independent Poisson random variables are Poisson random variables . . . . .	3
1.5	Exponential distribution $Exp(\lambda)$ : 独立随机事件的发生间隔/第一次发生事件的时间 . . . . .	3
1.6	Poisson process: A sequence of arrivals in continuous time with rate $\lambda$ . . . . .	4
1.6.1	Definition . . . . .	4
1.6.2	$T_j$ : time of $j^{th}$ arrival . . . . .	4
1.6.3	Theorem (Conditional counts): $N(t_1) N(t_2) = n \sim Bin(n, \frac{t_1}{t_2})$ . . . . .	5
<b>2</b>	<b>Limit Theorems</b>	<b>5</b>
2.1	Law of Large Numbers (LLN) . . . . .	5
2.2	Differences between <u>convergence in probability</u> (wLLN) and <u>wp1(a.s.)</u> (sLLN) . . . . .	6
2.3	Central Limit Theorem (CLT) . . . . .	6

# 1 Distribution

## 1.1 Bernoulli Distribution Bernoulli( $\pi$ ): 事件发生概率为 $\pi$

Assume  $n$  independent binary (taking values 0 or 1) observations arising from independent and identical trials:  $y_1, y_2, \dots, y_n$  such that:

$$P(Y_i = 1) = \pi \quad \text{and} \quad P(Y_i = 0) = 1 - \pi$$

Random variables  $Y_i$  are normally called Bernoulli trials.

$$Y_i \sim \text{Bernoulli}(\pi)$$

$$p(y) = \begin{cases} \pi & y = 1 \\ 1 - \pi & y = 0 \end{cases}$$

$$E(Y_i) = \pi, \text{Var}(Y_i) = \pi(1 - \pi)$$

## 1.2 Binomial distribution $\text{bin}(n, \pi)$ : $n$ 次 Bernoulli

The random variable  $Y = \sum_{i=1}^n Y_i$  has the Binomial distribution with index  $n$  and parameter  $\pi$  denoted as  $Y \sim \text{bin}(n, \pi)$ . Mass probability function for  $Y$ :

$$P(y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y} \quad y = 0, 1, 2, \dots, n$$

with  $\binom{n}{y} = n!/[y!(n-y)!]$

Mean and Variance:

$$E(Y) = \mu = n\pi \quad \text{var}(Y) = \sigma^2 = n\pi(1 - \pi)$$

Skewness:

$$E(Y - \mu)^3 / \sigma^3 = (1 - 2\pi) / \sqrt{n\pi(1 - \pi)}$$

If the independence assumption is violated, the Binomial distribution does not apply.

$$\frac{Y - n\pi}{\sqrt{n\pi(1 - \pi)}} \xrightarrow{d} N(0, 1)$$

(Normal approximation)

## 1.3 Multinomial Distribution

## 1.4 Poisson Distribution $Pois(\lambda)$ : 单位时间发生 $k$ 次事件的概率

$\lambda$ : 单位时间发生该时间的平均次数

$$\Pr(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, 3, \dots$$

$$E(X) = Var(X) = \lambda$$

推导:

我们考虑一段时间 (讲单位时间微分成  $n$  等分,  $n \rightarrow \infty$ ), 每一刻 (瞬间) 都有一个 event may occur, which follows binomial distribution  $B(n, p)$ . where  $n \rightarrow \infty, p \rightarrow 0$ ;  $\lambda = n \cdot p$  is the expected number of events in this period of time.

现在我们考虑发生  $k$  次 event 的概率:

$$\begin{aligned} \Pr(X = k) &= \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lim_{n \rightarrow \infty} \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \lim_{n \rightarrow \infty} \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k e^{-\lambda} \\ &= \frac{\lambda^k e^{-\lambda}}{k!} \lim_{n \rightarrow \infty} \frac{n!}{(n-k)!n^k} \\ &= \frac{\lambda^k e^{-\lambda}}{k!} \lim_{n \rightarrow \infty} \frac{n}{n} \frac{n-1}{n} \dots \frac{n-k+1}{n} \\ &= \frac{\lambda^k e^{-\lambda}}{k!} \end{aligned}$$

#### 1.4.1 Sums of independent Poisson random variables are Poisson random variables

$X \sim Pois(\lambda_1), Y \sim Pois(\lambda_2)$  are two independent Poisson random variables, then  $Z = X + Y$  also follow Poisson distribution, and the parameter is the sum of  $X$ 's and  $Y$ 's

$$Z \sim Pois(\lambda_1 + \lambda_2)$$

Then,

$$Z = X_1 + X_2 + \dots + X_n \sim Pois(\lambda_1 + \lambda_2 + \dots + \lambda_n)$$

#### 1.5 Exponential distribution $Exp(\lambda)$ : 独立随机事件的发生间隔/第一次发生事件的时间

$\lambda$ : 单位时间发生该时间的平均次数

随机变量  $X$  服从参数为  $\lambda$  或  $\beta$  的指数分布, 则记作

$$X \sim Exp(\lambda) \text{ or } X \sim Exp(\beta)$$

两者意义相同, 只是  $\lambda$  与  $\beta$  互为倒数关系.

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

$$f(x; \beta) = \begin{cases} \frac{1}{\beta} e^{-\frac{1}{\beta} x} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

累积分布函数为：

$$F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

其中  $\lambda > 0$  是分布的参数，即每单位时间发生该事件的次数； $\beta > 0$  为尺度参数，即该事件在每单位时间内的发生率。两者常被称为率参数 (rate parameter)。指数分布的区间是  $[0, \infty)$ 。

$\mathbb{E}(X) = \frac{1}{\lambda}$ ：预期事件的发生间隔； $Var(X) = \frac{1}{\lambda^2}$

$$\mathbb{E}(X) = \frac{1}{\lambda}; \quad Var(X) = \frac{1}{\lambda^2}$$

Memorylessness:  $\Pr(T > s + t \mid T > s) = \Pr(T > t)$

$$\begin{aligned} \Pr(T > s + t \mid T > s) &= \frac{\Pr(T > s + t \text{ and } T > s)}{\Pr(T > s)} \\ &= \frac{\Pr(T > s + t)}{\Pr(T > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\ &= e^{-\lambda t} \\ &= \Pr(T > t) \end{aligned}$$

推导：

我们考虑一段时间（讲单位时间微分成  $n$  等分， $n \rightarrow \infty$ ），每一刻（瞬间）都有一个 event may occur, which follows binomial distribution  $B(n, p)$ . where  $n \rightarrow \infty, p \rightarrow 0$ ;  $\lambda = n \cdot p$  is the expected number of events in this period of time. （与 Poisson 设定相同）

CDF: 现在我们考虑第一次发生 event 的时间大于  $x$  的概率：

$$1 - F(x; \lambda) = \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{nx} = e^{-\lambda x} \Rightarrow F(x; \lambda) = 1 - e^{-\lambda x}$$

PDF:

$$f(x; \lambda) = \frac{\partial F(x; \lambda)}{\partial x} = \lambda e^{-\lambda x}$$

## 1.6 Poisson process: A sequence of arrivals in continuous time with rate $\lambda$

### 1.6.1 Definition

$N(t) \sim Pois(\lambda t)$ : Number of arrivals in length  $t$  follows Poisson distribution

$$\begin{aligned} N(t) &\sim Pois(\lambda t) \\ \Pr(N(t) = k) &= \frac{(\lambda t)^k e^{-\lambda t}}{k!} \end{aligned}$$

The number of arrivals in disjoint time intervals are independent.

### 1.6.2 $T_j$ : time of $j^{th}$ arrival

$T_1 > t$  is same as  $N(t) = 0$ :  $P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t}$   
 $\Rightarrow T_1 \sim Expo(\lambda) \Rightarrow T_j - T_{j-1} \sim Expo(\lambda); T_j \sim Gamma(j, \lambda)$

**1.6.3 Theorem (Conditional counts):**  $N(t_1)|N(t_2) = n \sim \text{Bin}(n, \frac{t_1}{t_2})$

可以理解为  $n$  个点散落在  $(0, t_2]$  上的概率每处均等  $= \frac{1}{t_2}$ ; 所以散落在  $(0, t_1]$  上的概率为  $\frac{t_1}{t_2}$

## 2 Limit Theorems

### 2.1 Law of Large Numbers (LLN)

Describe the behavior of the sample mean of i.i.d. as the sample size grows.

$x_1, x_2, \dots, x_n$  i.i.d. with some distribution.  $\mu < \infty, \sigma^2 < \infty, \bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$ .

**Theorem 1** (Weak Law of Large Numbers (wLLN)).

*The weak law of large numbers (also called Khinchin's law) states that the sample average converges in probability towards the expected value.*

$$\bar{X}_n \xrightarrow{P} \mu \quad \text{when } n \rightarrow \infty.$$

That is, for any positive number  $\varepsilon$ ,

$$\lim_{n \rightarrow \infty} \Pr(|\bar{X}_n - \mu| < \varepsilon) = 1.$$

证明.

Proof: by Chebychev's inequality.

$$\begin{aligned} P(|\bar{x} - \mu| \geq \varepsilon) &\leq \frac{\sigma^2}{n\varepsilon^2} \quad (\text{Var } \bar{x} = \frac{\sigma^2}{n}) \\ \lim_{n \rightarrow \infty} \frac{\sigma^2}{n\varepsilon^2} &= 0 \\ \Rightarrow \lim_{n \rightarrow \infty} P(|\bar{x} - \mu| > \varepsilon) &\text{ also converges to } 0. \end{aligned}$$

□

**Theorem 2** (Strong Law of Large Numbers (sLLN)).

*With probability 1 (wp1) or almost surely (as).*

$$\bar{X}_n \xrightarrow{a.s.} \mu \quad \text{when } n \rightarrow \infty.$$

That is,

$$\Pr\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1.$$

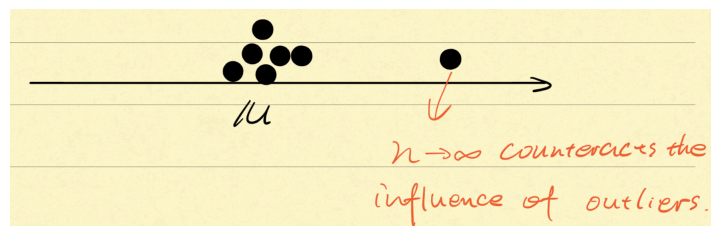


图 1: convergence in probability

## 2.2 Differences between convergence in probability (wLLN) and wp1(a.s.) (sLLN)

a) Weak Law of Large Numbers (wLLN)

$$P(|\bar{x} - \mu| \geq \varepsilon) \rightarrow 0 \text{ as } n \rightarrow +\infty, \forall \varepsilon > 0$$

b) Strong Law of Large Numbers (sLLN)

$$P(|\bar{x} - \mu| \geq \varepsilon \text{ as } n \rightarrow +\infty) = 0, \forall \varepsilon > 0$$

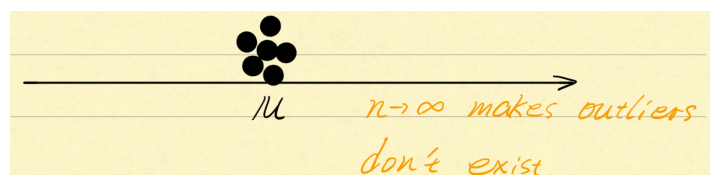


图 2: wp1(a.s.)

## 2.3 Central Limit Theorem (CLT)

**Theorem 3** (Central Limit Theorem (CLT)).

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{D} N(0, 1) \text{ when } n \rightarrow \infty$$

$Z$  converges in distribution to  $N(0, 1)$  as  $n \rightarrow \infty$

(converges in distribution:  $P(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq a) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{x^2}{2}} dx$ )

证明. Prove the situation of  $\mu = 0, \sigma^2 = 1$ , we can use linear transformations to get other situations.

Moment-generating function(MGF) of  $X_i$ :  $M_0(t) = E(e^{tX_i})$ .

$$M_0(0) = 1, M'_0(0) = EX_i = 0, M''_0(0) = EX_i^2 = 1$$

Moment-generating function(MGF) of  $\sqrt{n}\bar{X}$ :

$$\begin{aligned} M_1(t) &= Ee^{t\sqrt{n}\bar{X}} = Ee^{t\frac{\sum_{i=1}^n X_i}{\sqrt{n}}} \\ &= Ee^{t\frac{X_1}{\sqrt{n}}} \cdot Ee^{t\frac{X_2}{\sqrt{n}}} \dots Ee^{t\frac{X_n}{\sqrt{n}}} \\ &= [M_0(\frac{t}{\sqrt{n}})]^n \end{aligned}$$

$$\begin{aligned} \lim_{n \rightarrow \infty} \log M_1(t) &= \lim_{n \rightarrow \infty} n \log M_0(\frac{t}{\sqrt{n}}) \\ &\quad (\text{let } y = \frac{1}{\sqrt{n}}) \\ &= \lim_{y=0} \frac{\log M_0(yt)}{y^2} \\ &\quad (\text{L'Hôpital's rule}) \\ &= \lim_{y=0} \frac{tM_0'(yt)}{2yM_0(yt)} \\ &\quad (\text{L'Hôpital's rule}) \\ &= \lim_{y=0} \frac{t^2 M_0''(yt)}{2M_0(yt) + 2ytM_0'(yt)} \\ &= \frac{t^2}{2} \end{aligned}$$

As we know the Moment-generating function(MGF) of  $Z \sim N(0, 1)$  is  $M_Z(t) = \frac{t^2}{2}$ .  
Hence,  $M_1(t) = M_Z(t)$  i.e.  $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{D} N(0, 1)$  as  $n \rightarrow \infty$  □