Probability

Wenxiao Yang[*]

[*]Department of Mathematics, University of Illinois at Urbana-Champaign

2021

目录

1 Distribution

1.1 Bernoulli Distribution Bernoulli(π): 事件发生概率为 π

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

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

Random variables Y_i are normally called Bernoulli trials.

$$Y_i \sim \text{Bernoulli}(\pi)$$
$$p(y)=\begin{cases}\pi & y=1 \\ 1-\pi & y=0\end{cases}$$
$$E(Y_i)=\pi, Var(Y_i)=\pi(1-\pi)$$

1.2 Binomial distribution $bin(n,\pi)$: n 次 Bernoulli

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

$$P(y)=\left(\begin{array}{l}n \\ y\end{array}\right) \pi^y(1-\pi)^{n-y} \quad y=0,1,2, \ldots, n$$

with $\left(\begin{array}{l}n \\ y\end{array}\right)=n!/[y!(n-y)!]$
Mean and Variance:

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

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.

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

(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

λ: 单位时间发生该时间的平均次数

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

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

推导:

我们考虑一段时间 (讲单位时间微分成 n 等分, $n \to \infty$)，每一刻 (瞬间) 都有一个 event may occur, which follows binomial distribution $B(n,p)$. where $n \to \infty, p \to 0$; $\lambda = n \cdot p$ 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:

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

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

λ：单位时间发生该时间的平均次数

随机变量 X 服从参数为 λ 或 β 的指数分布，则记作

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

两者意义相同，只是 λ 与 β 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；$Var(X) = \frac{1}{\lambda^2}$

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

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, $n \to \infty$)，每一刻 (瞬间) 都有一个 event may occur, which follows binomial distribution $B(n, p)$. where $n \to \infty, p \to 0$; $\lambda = n \cdot p$ 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:

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

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 λ

1.6.1 Definition

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

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

The number of arrivals in disjoint time invervals 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): $N(t_1)|N(t_2) = n \sim Bin(n, \frac{t_1}{t_2})$

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

Theorem 1 (Weak Law of Large Numbers (wLLN))**.**
The weak law of large numbers (also called Khinchin's law) states that the sample average <u>*converges in probability*</u> *towards the expected value.*

$$\overline{X}_n \xrightarrow{P} \mu \qquad \textit{when } n \to \infty.$$

That is, for any positive number ε*,*

$$\lim_{n\to\infty} \Pr\big(\,|\overline{X}_n - \mu| < \varepsilon\,\big) = 1.$$

证明.
Proof: by Chebychev's inequality.

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

□

Theorem 2 (Strong Law of Large Numbers (sLLN))**.**
With probability 1 (wp1) or almost surely (as).

$$\overline{X}_n \xrightarrow{a.s.} \mu \qquad \textit{when } n \to \infty.$$

That is,

$$\Pr\Big(\lim_{n\to\infty} \overline{X}_n = \mu\Big) = 1.$$

2.2 Differences between <u>convergence in probability</u> (wLLN) and <u>wp1(a.s.)</u> (sLLN)

a) Weak Law of Large Numbers (wLLN)

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

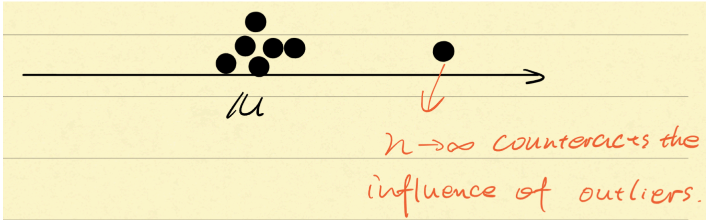

图 1: convergence in probability

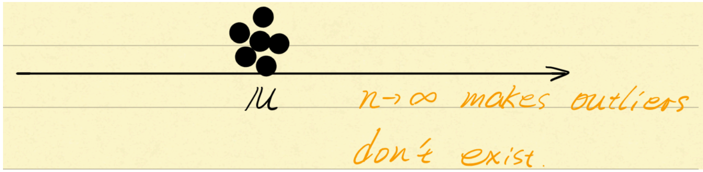

图 2: wp1(a.s.)

b) Strong Law of Large Numbers (sLLN)

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

2.3 Central Limit Theorem (CLT)

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

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

Z converges in distribution to $N(0,1)$ as $n \to \infty$
(converges in distribution: $P(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq a) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}} dx$)

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 $\sqrt{n}\overline{X}$:

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

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

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. $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{D} N(0,1)$ as $n \to \infty$ $\square$