第一章 随机变量的预备知识

Roman Vershynin

1.1 节介绍随机变量的期望, 方差和矩. 一些经典等式可以在 1.2 节中找到. 1.3 节回顾两个基本的概率极限理论: 大数定律与中心极限定理.

1.1 随机变量的数字特征

在概率论的基础课程中, 我们学习了随机变量 X 的两个最重要的数字特征: 期望 (也称均值) 和方差. 在本书中将它们表示为 $\mathbb{E}X$ 和 $Var\,(X) = \mathbb{E}(X - \mathbb{E}X)^2$.

我们回顾一下描述概率分布的其它经典数值特征和函数. X 的矩母函数被定义为 $M_X\,(t) = \mathbb{E}e^{tX}$, $t \in \mathbb{R}$, 对于 $p > 0$, X 的 p 阶矩记为 $\mathbb{E}X^p$, 且 p 阶绝对矩记为 $\mathbb{E}|X|^p$.

p 阶绝对矩的 p 次根是有用的, 这引出了随机变量的 L^p 范数的概念:

$$\|X\|_{L^p} = (\mathbb{E}|X|^p)^{\frac{1}{p}},\ p \in (0, \infty) \tag{1.1}$$

通过 $|X|$ 的本性上确界, 这个定义可以扩展到 $p = \infty$ 上去, 即

$$\|X\|_{L^\infty} = ess\ \sup|X| \tag{1.2}$$

关于上式的合理性, 可以参考 Thm 1.8.

对于给定的 p 和概率空间 (Ω, Σ, p), 经典向量空间 $L^p = L^p\,(\Omega, \Sigma, p)$ 有定义在 Ω 上的所有具有有限

L^p 范数的随机变量 X 构成, 即

$$L^p = \{X : \|X\|_{L^p} < \infty\} \tag{1.3}$$

如果 $p \in [1, \infty)$, 则 $\|X\|_{L^p}$ 是一个范数, L^p 是一个 Banach 空间. 这一事实源于 Minkowski 不等式 (Eq. 1.13). 对于 $p < 1$, 由于三角不等式不成立, 因此 $\|X\|_{L^p}$ 不是范数, 例如下面的例子,

Example 1.1 ($p = \frac{1}{2}$).

$$\begin{aligned} x &= \begin{bmatrix} 0.8147 & 0.9058 & 0.1270 & 0.9134 & 0.6324 \end{bmatrix}^T \\ y &= \begin{bmatrix} 0.0975 & 0.2785 & 0.5469 & 0.9575 & 0.9649 \end{bmatrix}^T \end{aligned} \tag{1.4}$$

but

$$\|x + y\|_{L^{\frac{1}{2}}} = \mathit{30.2049} \ > \ \mathit{28.2286} = \|x\|_{L^{\frac{1}{2}}} + \|y\|_{L^{\frac{1}{2}}} \tag{1.5}$$

L^p 空间中, 当指数 $p = 2$ 时, 因为 L^2 不仅是巴拿赫空间还是希尔伯特空间. L^2 上的内积和相应的范数由下式给出:

$$\langle X, Y \rangle_{L^2} = \mathbb{E}XY, \quad \|X\|_{L^2} = \left(\mathbb{E}|X|^2\right)^{\frac{1}{2}} \tag{1.6}$$

因而 X 的标准差可表示为

$$\|X - \mathbb{E}X\|_{L^2} = \sqrt{Var\,(X)} = \sigma\,(X) \tag{1.7}$$

同样, 可以将随机变量 X 和 Y 的协方差矩阵表示为:

$$\operatorname{cov}(X, Y) = \mathbb{E}\left((X - \mathbb{E}X)\,(Y - EY)\right) = \langle X - \mathbb{E}X, Y - \mathbb{E}Y \rangle_{L^2} \tag{1.8}$$

Remark. 随机变量中的几何学. 当我们将随机变量视为 L^2 空间 *(*希尔伯特空间*)* 中的向量时, 等式 *1.8* 给出了协方差的几何解释*:* 向量 $X - \mathbb{E}X$ 和 $Y - \mathbb{E}Y$ 越相关, 它们的内积和协方差越大.

Banach 空间 (a complete normed vector space), Hilbert 空间 (a complete inner product vector space).

1.2 一些经典不等式

Theorem 1.1 (Jensen's Inequality 1906)**.** 对于任意随机变量 X 和凸函数 $\phi: \mathbb{R} \to \mathbb{R}$, 我们有

$$\phi(\mathbb{E}X) \leqslant \mathbb{E}\phi(X) \tag{1.9}$$

证明. Assume that Jensen's inequality holds for distributions on n points, then

$$\begin{aligned}
\phi\left(\sum_{i=1}^{n+1}\lambda_i x_i\right) &= \phi\left(\sum_{i=1}^{n}\lambda_i x_i + \lambda_{n+1}x_{n+1}\right) \quad \textit{where} \sum_{i=1}^{n+1}\lambda_i = 1 \\
&= \phi\left(\sum_{i=1}^{n}(1-\lambda_{n+1})\frac{1}{1-\lambda_{n+1}}\lambda_i x_i + \lambda_{n+1}x_{n+1}\right) \\
&= \phi\left((1-\lambda_{n+1})\sum_{i=1}^{n}\frac{1}{1-\lambda_{n+1}}\lambda_i x_i + \lambda_{n+1}x_{n+1}\right) \\
&= (1-\lambda_{n+1})\,\phi\left(\sum_{i=1}^{n}\frac{1}{1-\lambda_{n+1}}\lambda_i x_i\right) + \lambda_{n+1}\phi(x_{n+1}) \quad \left(\sum_{i=1}^{n}\frac{\lambda_i}{1-\lambda_{n+1}} = 1\right) \\
&\le \sum_{i=1}^{n}\lambda_i\phi(x_i) + \lambda_{n+1}\phi(x_{n+1}) \\
&= \sum_{i=1}^{n+1}\lambda_i\phi(x_i)
\end{aligned} \tag{1.10}$$

□

Corollary 1.1. $\|X\|_{L^p}$ 是关于 p 的增函数, *i.e.*

$$\|X\|_{L^p} \leqslant \|X\|_{L^q}, \quad \forall 0 \leqslant p \leqslant q < \infty \tag{1.11}$$

证明. By $\phi(x) = x^{\frac{q}{p}}$ is convex, then

$$\|X\|_{L^p} = (\mathbb{E}|X|^p)^{\frac{1}{p}} = \left((\mathbb{E}|X|^p)^{\frac{q}{p}}\right)^{\frac{1}{q}} \leqslant \left[\mathbb{E}\left((|X|^p)^{\frac{q}{p}}\right)\right]^{\frac{1}{q}} = (\mathbb{E}(|X|^q))^{\frac{1}{q}} = \|X\|_{L^q} \tag{1.12}$$

□

Theorem 1.2 (Minkowski Inequality)**.** 对于任意 $p \in [1, \infty)$ 和任意随机变量 $X, Y \in L^p$, 我们有

$$\|X + Y\|_{L^p} \leqslant \|X\|_{L^p} + \|Y\|_{L^p} \tag{1.13}$$

可以将 1.13 看作是三角不等式, 其中当 $p \in [1, \infty)$ 时, $\|\cdot\|_{L^p}$ 是范数.

Theorem 1.3 (Hölder Inequality)**.** 如果 $p, q \in [1, \infty]$ 是共轭指数, *i.e.* $\frac{1}{p} + \frac{1}{q} = 1$, 则随机变量 $X \in L^p$ 和 $Y \in L^q$ 满足

$$|\mathbb{E}XY| \leqslant \|X\|_{L^p} \|Y\|_{L^q} \tag{1.14}$$

Corollary 1.2 (Cauchy-Schwarz Inequality)**.** 对任随机变量 $X, Y \in L^2$, 我们有

$$|\mathbb{E}XY| \leqslant \|X\|_{L^2} \|Y\|_{L^2} \tag{1.15}$$

You can see http://www.statslab.cam.ac.uk/~james/Lectures/pm.pdf

由概率论的基本概念可知, 随机变量 X 的分布直观反映了 X 所取值的概率信息. 更确切地说: X 的分布由 X 的累积分布函数 (Cumulative Distribution Function, CDF) 所确定, X 的累积分布函数定义为:

$$F_x(t) = \mathbb{P}\{X \leqslant t\}, \, t \in \mathbb{R} \tag{1.16}$$

使用随机变量的尾分布通常更方便, 即

$$\mathbb{P}\{X > t\} = 1 - F_x(t) \tag{1.17}$$

尾分布与随机变量的期望 (更一般地说, 矩) 之间存在着重要的关系. 下列等式通常用于界定尾分布的期望.

Lemma 1.2.1 (积分等式)**.** 设 X 为一个非负随机变量, 则有

$$\mathbb{E}X = \int_0^{+\infty} \mathbb{P}(X > t)\, dt \tag{1.18}$$

该等式的两端是同时有限或无限的.

证明. 可以通过如下等式表示任何非负实数 x,

$$x = \int_0^x 1 dt = \int_0^\infty 1_{(t<x)} dt \tag{1.19}$$

用随机变量 X 替换 x, 并对等式两边取期望, 得

$$\mathbb{E}X = \mathbb{E}\int_0^\infty 1_{(t<x)} dt \underset{by\ Fubini-Tonelli\ thm}{=} \int_0^\infty \mathbb{E}1_{(t<x)} dt = \int_0^\infty \mathbb{P}\{t < X\} dt \tag{1.20}$$

□

Homework 1.1 (积分不等式得推广)**.** *证明引理 1.2.1 的推广, 即求证引理 1.2.1 对任何随机变量 X (不一定非负) 都成立, 即*

$$\mathbb{E}X = \int_0^{\infty} \mathbb{P}\{X > t\}dt - \int_{-\infty}^{0} \mathbb{P}\{X < t\}dt \tag{1.21}$$

Homework 1.2 (通过尾分布计算 p 阶矩)**.** *设 X 是随机变量, $p \in (0, \infty)$. 求证:*

$$\mathbb{E}|X|^p = \int_0^{\infty} pt^{p-t}\mathbb{P}\{|X| > t\}\, dt \tag{1.22}$$

在右端有限时恒成立.

Theorem 1.4 (Markov 不等式)**.** *对于任何非负随机变量 X 和 $t > 0$, 有*

$$\mathbb{P}\{X \geqslant t\} \leqslant \frac{\mathbb{E}X}{t} \tag{1.23}$$

证明. 对任意 $t > 0$, 我们可以通过示性函数

$$x = x1_{(x \geqslant t)} + x1_{(x<t)} \tag{1.24}$$

来表示任意实数 x. 用随机变量 X 替换 x, 并对等式两边取期望, 则有

$$\begin{aligned} \mathbb{E}X &= \mathbb{E}X1_{(X \geqslant t)} + \mathbb{E}X1_{(X<t)} \\ &\geq \mathbb{E}t1_{(X \geqslant t)} + 0 \\ &= t\mathbb{P}\left(X \geqslant t\right) \end{aligned} \tag{1.25}$$

将不等式两边同时除以 t, 结论成立, 证毕. □

众所周知, Chebyshev 不等式是 Markov 不等式的特殊情况, 它给出了比对 t 更好的依赖性, 即关于 t 是二次依赖的, 且它不是控制单侧尾部, 而是量化了 X 关于其均值的集中度。

Corollary 1.3 (Chebyshev 不等式)**.** *设 X 是一个随机变量, 其均值为 μ, 方差为 σ^2, 则对任意 $t > 0$, 有*

$$\mathbb{P}\{|X - \mu| \geqslant t\} \leqslant \frac{\sigma^2}{t^2} \tag{1.26}$$

Homework 1.3. *通过将 $|X - \mu| \geqslant t$ 两端平方, 并应用 Markov 不等式推导出 Chebyshev 不等式.*

1.3 极限理论

独立随机变量之和的研究是经典概率论的核心部分. 回忆一下, 对于任意独立的随机变量 $X_1, X_2, ..., X_N$, 有

$$Var\left(X_1 + \cdots + X_N\right) = Var\left(X_1\right) + \cdots + Var\left(X_N\right) \tag{1.27}$$

Remark. *content...*

此外, 如果 X_i 是同分布的, 其均值为 μ, 方差为 σ, 将等式两端同时除以 N, 得

$$Var\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{\sigma^2}{N} \tag{1.28}$$

因此, 对于样本 $\{X_1, X_2, ..., X_N\}$, 当 $N \to \infty$ 时, 其样本均值 $\frac{1}{N}\sum\limits_{i=1}^{N} X_i$ 的方差趋于零. 这表明, 当样本量 N 足够大时, 其样本的均值最终将趋于总体的均值. 作为概率论中最重要的结果之一的大数定律恰恰说明了这一点.

Theorem 1.5 (强大数定律)**.** 设随机变量 $X_1, X_2, ...,$ 是独立同分布 *(i.i.d)* 的序列, 其均值为 μ , 考虑随机变量之和 $S_N = X_1 + \cdots + X_N$, 则当 $N \to \infty$ 时, 有

$$\frac{S_N}{N} \overset{a.e.}{\to} \mu \tag{1.29}$$

下面的中心极限定理进一步阐明了极限理论, 它将 X_i 的和的分布经适当缩放转化为正态分布 (也称作高斯分布). 回忆一下标准正态分布, 表示为 $N\left(0, 1\right)$, 其密度函数为

$$f\left(x\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \ \ x \in \mathbb{R} \tag{1.30}$$

Theorem 1.6 (Lindeberg-Lévy 中心极限定理)**.** 设随机变量 $X_1, X_2, ...$ 是独立同分布的序列, 其均值为 μ, 方差为 σ^2. 考虑随机变量之和 $S_N = X_1 + \cdots + X_N$, 并对其进行标准化以获得具有零均值和单位方差的随机变量, 即

$$Z_N : \frac{S_N - \mathbb{E}S_N}{\sqrt{Var\left(S_N\right)}} = \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}\left(X_i - \mu\right) \tag{1.31}$$

则当 $N \to \infty$ 时, 有

$$Z_N \overset{\text{依分布}}{\to} N(0,1) \tag{1.32}$$

依分布收敛意味着随机变量序列的部分和经标准化后其分布趋于标准正态分布. 我们也可以用随机变量的尾分布表示, 即对任意 $t \in \mathbb{R}$, 当 $N \to \infty$ 时, 有

$$\mathbb{P}\{Z_N \geqslant t\} \to \mathbb{P}\{g \geqslant t\} = \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-\frac{x^2}{2}} dx \tag{1.33}$$

其中 $g \sim N(0,1)$ 是服从标准正态分布的随机变量.

Homework 1.4. 设随机变量 $X_1, X_2, \ldots$ 是独立同分布的序列, 其均值为 μ, 方差有限, 求证当 $N \to \infty$ 时, 有

$$\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^{N} X_i - \mu\right| = O\left(\frac{1}{\sqrt{N}}\right) \tag{1.34}$$

Remark. *https://en.wikipedia.org/wiki/Big_O_notation*

中心极限定理的一个典型的特殊情况为: 当 X_i 服从参数为 $p \in (0,1)$ 的伯努利分布时, 记为 $X_i \sim Ber(p)$, 该分布表示 X_i 取值 1 和 0 的概率分别为 p 和 $1-p$, 则 $\mathbb{E}X_i = p$, $Var(X_i) = p(1-p)$. 随机变量之和 $S_N \triangleq X_1 + \cdots + X_N$ 被称为服从二项分布 $\text{Binon}(N,p)$. 则由中心极限定理 1.6 知, 当 $N \to \infty$ 时, 有

$$\frac{S_N - Np}{\sqrt{Np(1-p)}} \overset{\text{依分布}}{\to} N(0,1) \tag{1.35}$$

中心极限定理的这种特殊情况被称为 de Moivre-Laplace (棣莫弗-拉普拉斯) 定理.

现在假设 $X_i \sim Ber(p_i)$, 当 $N \to \infty$ 时, 参数 P_i 趋于零, 且随机变量之和 S_N 有均值 $O(1)$, 而不是与 N 成比例. 中心极限定理在该情况下不成立. 我们即将陈述的另一个结果表明, S_N 仍然会收敛, 但是收敛到泊松分布而不是正态分布.

回忆一下, 随机变量 Z 服从参数 λ 的泊松分布, 表示为 $Z \sim Pois(\lambda)$, 如果 Z 取值为 $\{0,1,2,\ldots\}$, 且 $\mathbb{P}(Z=K) = e^{-\lambda}\frac{\lambda^k}{k!}$, $k = 0,1,2,\ldots$

Theorem 1.7 (泊松极限定理)**.** 设 $X_{N,i}$ $(1 \leqslant i \leqslant N)$ 是独立的随机变量, 且 $X_{N,i} \sim Ber(p_{N,i})$, 记 $S_N = \sum\limits_{i=1}^{N} X_{N,i}$. 如果 $N \to \infty$ 时, 有

$$\max_{i\leqslant N} p_{N,i} \to 0,\ \mathbb{E}S_N = \sum_{i=1}^{N} p_{N,i} \to \lambda < \infty \tag{1.36}$$

那么，当 $N \to \infty$ 时，有

$$S_N \overset{\text{依分布}}{\to} Pois\left(\lambda\right). \tag{1.37}$$

1.4 补充

本章介绍的内容包含在大多数的研究生概率论教科书中. 读者可以在参考文献 [1] section 1.7 & 2.4 和 [2] section 6 & 27 中找到 1.5 和 1.6 的证明.

Definition 1.1. *The L^p norm is formally defined as:*

$$\|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p} \tag{1.38}$$

The L^p norm has several cases that supposedly arise often in linear algebra, numerical analysis, and machine learning. The L_∞ norm:

$$\|x\|_\infty = \max_i |x_i| \tag{1.39}$$

Theorem 1.8. *We have that*

$$\lim_{p\to\infty} \|x\|_p = \|x\|_\infty \tag{1.40}$$

证明. We can do that

1. $p \geq 1$

$$\begin{aligned} \|x\|_\infty^p &= \max_i |x_i|^p \\ \|x\|_p^p &= \sum_i |x_i|^p \end{aligned} \tag{1.41}$$

$$\therefore \quad \|x\|_\infty \leqslant \|x\|_p \tag{1.42}$$

2. $p \geq 1$

$$\|x\|_p = \|x\|_\infty \cdot \frac{\left(\sum_i |x_i|^p\right)^{1/p}}{\|x\|_\infty} = \|x\|_\infty \cdot \left(\sum_i \left(\frac{|x_i|}{\|x\|_\infty}\right)^p\right)^{1/p} \leqslant \|x\|_\infty \cdot n^{1/p} \tag{1.43}$$

3. finally, by Eq. 1.42 and 1.43, we can get

$$\|x\|_\infty \leqslant \|x\|_p \leqslant \|x\|_\infty n^{1/p} \tag{1.44}$$

so, taking a limit as $p \to \infty$, we have

$$\|x\|_\infty \leqslant \lim_{p\to\infty} \|x\|_p \leqslant \|x\|_\infty \lim_{p\to\infty} n^{1/p} = \|x\|_\infty \tag{1.45}$$

$$\therefore \|x\|_\infty = \lim_{p\to\infty} \|x\|_p. \tag{1.46}$$

□

1.5 参考文献

[1] R. Durrett , Probablity: Theory and Examples, *Cambridge Series in Statistical and Probabilistic Mathematics*, 2010, Vol. 31.

[2] P. Billingsley , Probability and Measure, *Wiley Series in Probability and Mathematical Statistics*, 1995.