Math in Information Age

Course Review

Cheng Chen, Yican Sun

EECS, PKU

Outline

1. 概率复习
2. 作业
3. 课堂回顾

概率复习

概率的公理化定义

设 Ω 为任意的非空集合，叫做基本事件空间，也叫样本空间。事件 A 定义为 Ω 的一个子集。$\mathcal{F}$ 为一系列事件的集合，P 为定义在 F 上的函数。

若 $\mathcal{F}$ 满足:

- $\Omega \in \mathcal{F}$
- 若 $A \in \mathcal{F}$，则 $\Omega - A \in \mathcal{F}$
- 若 $A_n \in \mathcal{F}(n = 1, 2, 3, ...)$，则 $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$

且函数 P 满足:

- $P(\Omega) = 1$
- $P(\overline{A}) = 1 - P(A)$
- 若 $A_n \in \mathcal{F}$ 且两两不相交，则有 $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$

则称 $(\Omega, \mathcal{F}, P)$ 为概率空间。

随机变量的定义

注意到概率空间适用的范围非常广泛，甚至大部分基本事件不能用数学表述。为了将这些非数学语言进行量化，我们引入了随机变量的概念。

对于概率空间 $(\Omega, \mathcal{F}, P)$，$X = X(\omega)$ 是在 Ω 上有定义的实值函数。若对任意 x，$\{\omega | X(\omega) \leq x\} \in \mathcal{F}$，则称 X 为该概率空间上的随机变量。

[1]

定义了随机变量后，我们就可以把事件的概率迁移到随机变量取值的概率：$P(X = t) = P(\{\omega | X(\omega) = t\})$。

[1]实际上，解决实际问题时我们选取的事件集合 $\mathcal{F}$ 足够大，往往不注重 $\{\omega | X(\omega) \leq x\} \in \mathcal{F}$ 这个条件。

随机变量举例

小明第一次去买彩票。由于小明很穷，他决定这次买两种彩票：第一种彩票为猜测明天的天气是否是晴天，第二种彩票为猜测明天矿泉水会不会涨价。

小明分别投注了明天是晴天，明天矿泉水会涨价。但小明的心态很差：如果两次彩票都没中，他认为这是彩票公司坑人钱财，便再也不会买彩票；如果两次都中了，小明心满意足也就不再买彩票；只有一个猜中另一个猜不中时，小明才会被勾起兴趣，会再来买彩票。

已知明天晴天的概率是 p_1，矿泉水涨价的概率是 p_2，请问小明会再来买彩票的概率。[2]

[2]概率可能起源于赌博，本题仅沿用类似情景

随机变量

- 离散型随机变量只能取有限个值或可列无穷个值。比如:
  - 伯努利分布(两点分布): $P(X=1)=p, P(X=0)=1-p$
  - 泊松分布: $P(X=k)=\frac{1}{k!}\lambda^k e^{-\lambda}$
- 连续型随机变量满足:

$$\exists p(x), \forall a<b, P(a<X<b)=\int_a^b p(x)dx$$

  这里 $p(x)$ 被称为概率密度函数。比如:
  - 均匀分布: $p(x)=\begin{cases}\frac{1}{b-a} & a\leq x\leq b \\ 0 & else\end{cases}$
  - 正态分布: $p(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$
- 混合随机变量

随机变量的数学期望

随机变量的期望:

- 离散型随机变量: $E(X)=\sum_{k=1}^{\infty}x_kp_k$
- 连续性随机变量: $E(X)=\int_{-\infty}^{+\infty}xp(x)dx$

随机变量函数的数学期望: $\int_{-\infty}^{+\infty}f(x)p(x)dx$

关于随机变量的方差 $E(X-E(X))^2$，这本质上可以看作是一种随机变量函数的期望。但方差有一个比较好的性质:

$$E(X-E(X))^2=E(X^2)-2E(XE(X))+(E(X))^2=E(X^2)-E(X)^2$$

因此在计算方差时，往往从 $E(X^2)$ 入手。

随机变量的数学期望

设 X 是只取非负整数值的随机变量，且 $E(X)$ 存在，证明：

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k)$$

条件概率与独立性

条件概率公式：$P(A|B) = P(A \cap B)/P(B)$

对于两个事件 A, B，若 $P(A \cap B) = P(A)P(B)$，则称他们相互独立。

对于多个事件 $A_1, A_2, ...A_n$，若对于任意下标集合 $I = \{i_1, i_2, ..., i_k\}$，有 $P(\cap_{t=1}^{k} A_{i_t}) = \prod_{t=1}^{k} P(A_{i_t})$，则称这 n 个事件相互独立。

条件概率的重要公式

设事件组 $B_1, B_2, ...B_n$ 满足下列条件:

- $B_1, B_2, ...B_n$ 两两不相容，且 $P(B_i) > 0$
- $P(\cup_{i=1}^{n} B_i) = 1$

则对任何事件 A,

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i)\text{(全概公式)}$$

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}\text{(逆概公式)}$$

逆概公式举例

小明觉得买彩票跟运气成分太大，决定参加北京大学知识竞赛。知识竞赛的第一题是一道 10 个选项的常识单选题。场外评委发现小明从东北门入校，于是认为小明只有 0.01 的概率知道真正答案，而选手不知道答案时近似仍为是在瞎猜，有 $\frac{1}{10}$ 的概率猜对。

现在小明选出了正确答案。请问在小明选出正确答案的条件下，小明真正知道正确答案的概率是多少？

条件分布和联合分布

对于随机变量 X，Y，和条件概率的出发点一样，我们想要观察 X 和 Y 之间的关系，以连续型随机变量为例子，一般来说会从以下几个方面才来考虑：

- 联合分布：二维随机向量 (X,Y) 的分布，即二维概率密度 $p_{X,Y}(x,y)$，对应概率 $P(a<X<b,c<Y<d)=\iint_{a<x<b,c<y<d} p_{X,Y}(x,y)dxdy$
- 边缘分布：得到 $p_{X,Y}(x,y)$ 后，反过来推导 $p_X(x)=\int_{-\infty}^{+\infty} p_{X,Y}(x,y)dy$
- 条件分布： $p_{X|Y}(x|y)=\frac{p(x,y)}{p_Y(y)}$

类似事件的独立性，我们定义随机变量相互独立为：
$p_{X,Y}(x,y)=p_X(x)p_Y(y)$（连续型随机变量），
$P(X=x,Y=y)=P(X=x)P(Y=y)$（离散型随机变量）。

除了独立性，还有不相关性：

$$E(XY)=E(X)E(Y)$$

一些重要的离散分布

- 两点分布：$P(X=1)=p, P(X=0)=1-p$
- 二项分布：$P(X=k)=C_n^k p^k (1-p)^{n-k}$，记作 $X \sim B(n,p)$
  二项分布可以从两点分布而来。
  考虑 n 个服从上述两点分布的独立随机变量 $X_1, X_2, ...X_n$，令 $Y=\sum_{i=1}^n X_i$，则 $P(Y=k)=C_n^k p^k (1-p)^{n-k}$。
- 泊松分布：$P(X=k)=\frac{1}{k!}\lambda^k e^{-\lambda}$
  泊松分布可以看作是二项分布中 n 趋近于无穷的情况。

正态分布

正态分布 $N(\mu, \sigma)$ 为如下形式：

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

中心极限定理：设 $X_1, X_2, \dots$ 是独立同分布的随机变量序列，$\mu = E(X)$ 和 $\sigma^2 = var(X)$ 都存在，设 $S_n = \sum_{i=1}^{n} X_i$，则对一切 x，下列式子成立：

$$\lim_{n \to \infty} P(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

正态分布

正态分布 $X \sim N(\mu, \sigma)$:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

正态分布有如下性质:

- $E(X) = \mu, var(X) = \sigma^2$
- $X \sim N(\mu_1, \sigma_1), Y \sim N(\mu_2, \sigma_2)$，则 $X + Y \sim N(\mu_1 + \mu_2, \sigma_1 + \sigma_2)$
- 高维正态分布

$$p(\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}$$

概率相关的不等式

- Union Bound

$$P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$$

- Tail Bound: $x_1, x_2, ..., x_n$ 为参数为 p 的伯努利分布，$s = \sum_{i=1}^{n} x_i$，容易得到 $E(s) = np$。令 $m = np$，对于 s 有下列不等式成立：
  - $\forall \delta > 0, P(s > (1+\delta)m) < (\frac{e^{\delta}}{(1+\delta)^{1+\delta}})^m$
  - $Let\ 0 < \gamma \leq 1,\ P(s < (1-\gamma)m) < (\frac{e^{-\gamma}}{(1-\gamma)^{1-\gamma}})^m < e^{-\frac{\gamma^2 m}{2}}$
-

作业

课堂回顾

Markov Inequality

对于随机变量 $X \geq 0$

$$\Pr[X \geq k] \leq \frac{\mathbb{E}[X]}{k}$$

Proof: $\mathbb{E}[X] \leq \Pr[0 \leq X < k] \cdot 0 + \Pr[X \geq k] \cdot k$

tight example: 考虑一个随机变量, 取在 0 的概率是 $1 - \frac{1}{k}$, 取在 k 的概率是 $\frac{1}{k}$

概率不等式

Chebyshev Inequality

$$\Pr[|X-\mathbb{E}[X]|\geq a]\leq \frac{Var[X]}{a^2}$$

Proof: $\Pr[|X-\mathbb{E}[X]|\geq a]=\Pr[(X-\mathbb{E}[X])\geq a^2]\leq \frac{Var[X]}{a^2}$

tight example:(Exercise)

Law of Large Numbers

令 $x_1, x_2, \cdots, x_n$ 是随机变量 X 的 n 个样本. 那么:

$$\Pr[|\frac{1}{n}\sum_{i=1}^{n} x_i - \mathbb{E}[X] > \epsilon|] \leq \frac{Var[X]}{n\epsilon^2}$$

Proof: 考虑 $\mu_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, 那么 $\mathbb{E}[\mu_n] = \mathbb{E}[X]$, $Var[\mu_n] = \frac{1}{n}Var[X]$, 用 Chebyshev inequality 即可

高维空间的单位球

体积，面积公式

$$A(d) = \frac{2\pi^{d/2}}{\Gamma(d/2)}$$

$$V(d) = A(d)/d = \frac{2\pi^{d/2}}{d\Gamma(d/2)}$$

高维空间的单位球

- 体积集中在表面。
- $\frac{1}{d}$ 是一个转折点.
  - 几乎没有点与表面的距离是 $o(\frac{1}{d})$
  - 几乎没有点与表面的距离是 $\omega(\frac{1}{d})$

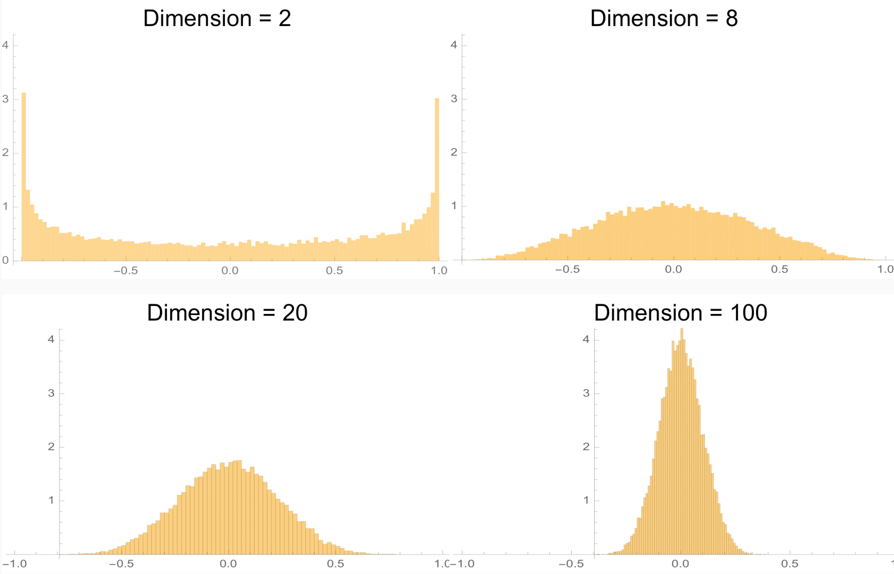
  Proof:

$$\frac{\mathrm{volumn}((1-\epsilon)S)}{\mathrm{volumn}(S)} = (1-\epsilon)^d \to \exp(-\frac{d}{\epsilon})$$

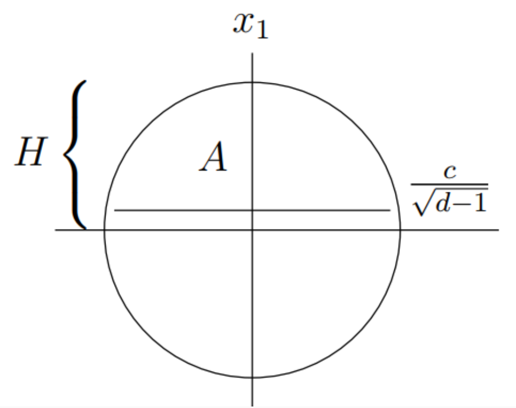
高维空间的单位球

- 体积集中在赤道上。

对于 $d \geq 3$ 和 $c \geq 1$, 至多只有 $\frac{2}{c}\exp(-\frac{c^2}{2})$ 的体积满足 $|x_1| > \frac{c}{\sqrt{d-1}}$


Dimension = 2
Dimension = 8
Dimension = 20
Dimension = 100
0
1
2
3
4
−1.0
−0.5
0.0
0.5
1.0

Proof

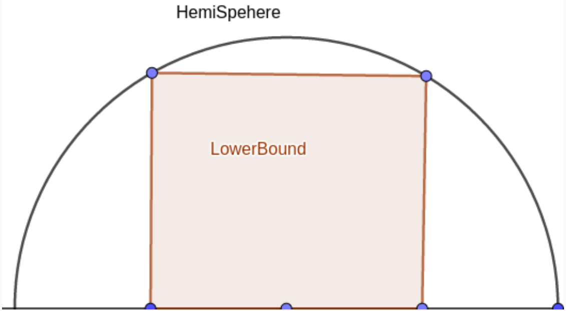


我们给出 A 的体积的上界，和 H 体积的下界。

A 体积的上界

$$\begin{aligned}
\mathrm{Volumn}(A) &= \int_{\frac{c}{\sqrt{d-1}}}^{1} (1-x^2)^{(d-1)/2} V(d-1)\mathrm{d}x \\
&\leq \int_{\frac{c}{\sqrt{d-1}}}^{1} \exp(-\frac{(d-1)x^2}{2}) V(d-1)\mathrm{d}x \\
&\leq \int_{\frac{c}{\sqrt{d-1}}}^{\infty} \exp(-\frac{(d-1)x^2}{2}) V(d-1)\mathrm{d}x \\
&\leq \int_{\frac{c}{\sqrt{d-1}}}^{\infty} \frac{x\sqrt{d-1}}{c} \exp(-\frac{(d-1)x^2}{2}) V(d-1)\mathrm{d}x \\
&= \frac{V(d-1)}{c\sqrt{d-1}} \exp(-c^2/2)
\end{aligned}$$

H 体积的下界



柱体高 $\frac{1}{\sqrt{d-1}}$, 半径 $\sqrt{1-\frac{1}{d-1}}$

因此体积是 $V(d-1)(1-\frac{1}{d-1})^{(d-1)/2}\frac{1}{\sqrt{d-1}} \geq \frac{V(d-1)}{2\sqrt{d-1}}(d \geq 3)$

球面是否分布在赤道附近？(Exercise)

球面与赤道

球面分布在赤道附近。

H 表面积的上界 $\frac{A(d-1)}{c\sqrt{d-2}}\exp(-c^2/2)$.

A 的表面积 $\frac{A(d)}{2}$.

比一比: $\frac{2A(d-1)}{A(d)c\sqrt{d-2}} = \frac{2\Gamma(\frac{d-1}{2})}{c\sqrt{\pi(d-2)}\Gamma(d/2)}\exp(-c^2/2)$

Theorem 2.8

令 $x_1, x_2, \ldots, x_n$ 是按照 n 个单位球内均匀分布取出来的向量。那么至少有 $1 - O(\frac{1}{n})$ 的概率，下面两件事情同时成立：

- $\forall i, \qquad |x_i| \geq 1 - \frac{2 \ln n}{d}$
- $\forall i, j, \quad |\langle x_i, x_j \rangle| \leq \frac{\sqrt{6 \ln n}}{\sqrt{d-1}}$

Proof:

$$\Pr[|x_i| \leq 1 - \frac{2 \ln n}{d}] \leq \exp(-2 \ln n) = \frac{1}{n^2}$$

$$\Pr[|\langle x_i, x_j \rangle| > \frac{\sqrt{6 \ln n}}{\sqrt{d-1}}] \leq \frac{2}{\sqrt{6 \ln n}} \exp(-\frac{6 \ln n}{2}) = O(n^{-3})$$

高维立方体 (Exercise)

- 高维立方体就是 $[0,1]^d$.
- 定义赤道为超平面 $\sum x_i = \frac{d}{2}$.
- 体积是否都分布在表面附近？
- 顶点是否都分布在赤道附近？
- 表面积是否都分布在赤道附近？