1 概率论基本概念

概率公理化：S 为样本空间，F 为 S 的某些子集组成的事件域。如果定义在 F 上的实值函数 P 满足。1. $\forall A\in F,P(A)\geq 0$；2. $P(S)=1$；3. $\forall A_1,A_2,\ldots\in F$ 且两两互斥，有 $P\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i)$，则称 P 为概率测度，(S,F,P) 为概率空间。

一般加法公式 $P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leq i<j\leq n}P(A_iA_j)+\cdots+(-1)^{n-1}P(A_1\cdots A_n)$

贝叶斯公式 $P(A|B)=P(B|A)\cdot\dfrac{P(A)}{P(B)}$

相互独立比两两独立**更强**。

Union Bound $P(\bigcup A_i)\leq\sum P(A_i)$，对 A_i 无要求。

证明存在可以通过证明其概率大于 0

2 离散随机变量

$P(X\geq E(X))>0,P(X\leq E(X))>0,\mathrm{Var}(aX+b)=a^2\mathrm{Var}(X)$

尾不等式/集中不等式：随机变量与期望的偏离。

Markov：对**非负** X，$E(X)>0,a>0$，有

$$P(X\geq a)\leq E(X)/a \quad P(X\geq aE(X))\leq 1/a$$

Chebyshev：对 $\sigma(X)>0,c>0$，

$$P(|X-E(X)|\geq c\cdot\sigma(X))\leq 1/c^2$$

$$P(|X-E(X)|\geq a)\leq \mathrm{Var}(X)/a^2$$

分布	分布列 $P(X=k)=$	期望	方差
$B(n,p)$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
$G(p)$	$p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$\pi(\lambda)$	$\lambda^k e^{-\lambda}/k!$	λ	λ
$NB(r,p)$	$\binom{k-1}{r-1}p^r(1-p)^{k-r}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

二项分布：n 次独立重复伯努利实验成功次数。**几何分布**：第一次成功的试验次数。**负二项分布**：第 r 次成功的次数。有 $G(p)=NB(1,p)$。**泊松分布**：单位时间内事件发生次数。几何分布具有**无记忆性**：$P(X>m+n|X>m)=P(X>n)$。$NB(r,p)$ 可以拆成 r 个独立的 $G(p)$ 之和。

3 连续随机变量

正态分布：$X\sim N(\mu,\sigma^2)$，$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$，$E(X)=\mu,\mathrm{Var}(X)=\sigma^2$。标准正态分布 $Z\sim N(0,1)$。

指数分布：$X\sim\mathrm{Exp}(\lambda)$，$f(x)=\lambda e^{-\lambda x}$，$F(x)=1-e^{-\lambda x}$，$E(X)=1/\lambda,\mathrm{Var}(X)=1/\lambda^2$。具有无记忆性 $P(X>s+t|X>s)=P(X>t)$。理解为泊松分布假设下，第一次事件发生的时刻。

伽马分布：$X\sim\Gamma(\alpha,\lambda)$，理解为泊松假设下第 α 次的时刻。

$$f(x)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x},x\geq 0$$

$E(X)=\alpha/\lambda$，$\mathrm{Var}(X)=\alpha/\lambda^2$。其中 $\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}\mathrm{d}x,\Gamma(\alpha+1)=\alpha\cdot\Gamma(\alpha),\Gamma(n+1)=n!,\Gamma(n+\frac{1}{2})=\frac{(2n)!}{4^n n!}\sqrt{\pi}$。具有可加性：$\Gamma(\alpha_1+\alpha_2,\lambda)=\Gamma(\alpha_1,\lambda)+\Gamma(\alpha_2,\lambda)$。$\alpha=1$ 时即指数分布。$\alpha=n/2,\lambda=1/2$ 时即 $\chi^2(n)$ 分布。

概率密度变换：$Y=g(X)$，g 单调且反函数 $h(y)$ 有连续导数，则 $f_Y(y)=f_X(h(y))\cdot|h'(y)|$。如果没法直接套这个公式的话可以从**分布函数**的定义出发进行变换。

4 多维离散随机变量

对于**独立**的 X,Y，有 $\mathrm{Var}(X\pm Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$

协方差：$\mathrm{Cov}(X,Y)=E((X-E(X))(Y-E(Y)))=E(XY)-E(X)E(Y)$。

$\mathrm{Var}(X_1+\cdots+X_n)=\sum_i\sum_j\mathrm{Cov}(X_i,X_j)$，

$\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y)$，

$\mathrm{Cov}(aX,bY)=ab\cdot\mathrm{Cov}(X,Y)$，

$\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$

条件期望：$E(X|Y=y)=\sum_i x_iP(X=x_i|Y=y)$，是关于 y 的函数。$E(X|Y)$ 是随机变量。

重期望公式：$E(E(X|Y))=E(X)$

5 多维连续随机变量

二维高斯：$X,Y\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，要求 $|\rho|<1$。

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2}+\frac{(y-\mu_2)^2}{\sigma_2^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right]$$

边际密度函数与 ρ 无关。$\rho=0$ 时独立。协方差 $\mathrm{Cov}(X,Y)=\rho\sigma_1\sigma_2$。

相关系数 $\mathrm{Corr}(X,Y)=\mathrm{Cov}(X,Y)/\sigma(X)\sigma(Y)$。相关系数/协方差大于 0 则正相关，小于 0 则负相关，等于 0 则不相关（但不一定独立）。相关系数等于 ± 1 代表 X,Y 呈严格线性关系。证明考虑标准化 $\tilde{X},\tilde{Y}$，然后通过 $\mathrm{Var}(\tilde{X}-\tilde{Y})=0$ 推导出 $P(\tilde{X}-\tilde{Y}=c)=1$。

5.1 概率密度变换

卷积公式：若 X,Y 独立，$Z=X+Y$，则

$$f_Z(z)=\int_{-\infty}^{\infty}f_X(x)f_Y(z-x)\mathrm{d}x=\int_{-\infty}^{\infty}f_X(z-y)f_Y(y)\mathrm{d}y$$

min/max：若 $X_1,\cdots,X_n$ 独立，则 $Y=\max\{X_i\}$ 的分布函数为 $F_Y(y)=\prod_{i=1}^{n}F_{X_i}(y)$，$Y=\min\{X_i\}$ 的分布函数为 $F_Y(y)=1-\prod_{i=1}^{n}(1-F_{X_i}(y))$。

换元：X,Y 的概率密度为 $f(x,y)$，函数 $u=u(x,y)$ 和 $v=v(x,y)$ 偏导连续且 $x=x(u,v),y=y(u,v)$ 为唯一反函数，则 $U=u(X,Y),V=v(X,Y)$ 的联合概率密度为

$$g(u,v)=f(x(u,v),y(u,v))\cdot|J|,J=\left|\frac{\partial(x,y)}{\partial(u,v)}\right|=\begin{vmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{vmatrix}$$

5.2 线性代数

对角化：$\boldsymbol{A}=\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}$。保行列式，平方的行列式和 trace 不变。实对称矩阵可对角化，不同特征值的特征向量正交。

正定：半正定矩阵 $\boldsymbol{A}$ 存在 $\boldsymbol{B}=\boldsymbol{A}^{1/2}$，$\boldsymbol{B}^\top\boldsymbol{B}=\boldsymbol{B}^2=\boldsymbol{A}$，且 B 不唯一。对于正定矩阵，这样的 $\boldsymbol{B}$ 可逆，$(\boldsymbol{B}^{-1})^2=\boldsymbol{A}^{-1}$。

协方差矩阵：对于随机变量 $\boldsymbol{X}=(X_1,\cdots,X_n)$，$\mathrm{Cov}(\boldsymbol{X})=E((\boldsymbol{X}-E(\boldsymbol{X}))(\boldsymbol{X}-E(\boldsymbol{X}))^\top)$ 为协方差矩阵，有

$$\mathrm{Cov}(\boldsymbol{X})=\begin{pmatrix}\mathrm{Var}(X_1) & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_n)\\ \mathrm{Cov}(X_2,X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2,X_n)\\ \vdots & \vdots & & \vdots\\ \mathrm{Cov}(X_n,X_1) & \mathrm{Cov}(X_n,X_2) & \cdots & \mathrm{Var}(X_n)\end{pmatrix}$$

其对称且半正定。

高斯：n 维高斯的联合密度函数，$\boldsymbol{B}$ 为协方差矩阵：

$$f(\boldsymbol{x})=\frac{1}{(2\pi)^{\frac{n}{2}}(\det\boldsymbol{B})^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\cdot(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

若 $\boldsymbol{X}\sim N(\boldsymbol{\mu},\boldsymbol{B})$，令 $\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}+\boldsymbol{b}$，且 $\boldsymbol{A}$ 行满秩，则 $Y\sim N(\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b},\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^\top)$

6 尾不等式、大数定律与中心极限定理

尾不等式：$P(X\geq k)$ 的上界。集中不等式：$P(|X-E(X)|\geq k)$ 的上界。

矩：$E(X^n)$ 称为 X 的 n 阶矩，$E((X-E(X))^n)$ 称为 X 的 n 阶中心矩。切比雪夫不等式的本质是对二阶中心矩使用 Markov。

矩生成函数：$M_X(t)=E(e^{tX})=\sum_{i\geq 0}\frac{t^i}{i!}E(X^i)$。所以求 k 阶矩可以求其封闭形式的 k 阶导然后令 $t=0$。

Chernoff Bound：求 k 阶中心矩然后用 Markov 得到的尾不等式通常较弱（没有真正用到 n 重伯努利实验的独立性）。对于任意 $t>0$，有

$$P(X\geq a)=P(e^{tX}\geq e^{ta})\leq\frac{E(e^{tX})}{e^{ta}}=e^{-ta}M_X(t)$$

对于任意 $t<0$ 有 $P(X\leq a)\leq M_X(t)\cdot e^{-ta}$ 通过调节 t 可得到更紧的上界。一般而言是**求导**找最小值。

Hoeffding 引理：若实数随机变量 $a\leq X\leq b$，则对任意实数 t 有

$$E(e^{t(X-E(X))})\leq\exp\left(\frac{t^2(b-a)^2}{8}\right)$$

Chernoff-Hoeffding 不等式：$X_1,\cdots,X_n$ 独立，且 $a_i\leq X_i\leq b_i$，令 $X=\sum_{i=1}^{n}X_i$，则对任意 $t>0$ 有

$$P(X\geq E(X)+t)\leq\exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2}\right)$$

$$P(X\leq E(X)-t)\leq\exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2}\right)$$

对于 $a\leq X_i\leq b$ 的情况，分母就是 $n(b-a)^2$。

大数定律的一般形式：对于随机变量 $\{X_n\}$，对于任意 $\epsilon>0$，若

$$\lim_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)\right|<\epsilon\right)=1$$

Markov 大数定律：$\frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n}X_i)\to 0$，则 $\{X_n\}$ 满足大数定律的一般形式。

辛钦大数定律：$\{X_n\}$ 独立同分布，且 $E(X_i)=\mu$，则 $\{X_n\}$ 满足大数定律的一般形式。对比 Markov，需要 iid，但不需要方差。

依概率收敛：随机变量序列 $\{X_n\}$ 依概率收敛于 X，记作 $X_n\xrightarrow{P}X$，如果对任意 $\epsilon>0$，有

$$\lim_{n\to\infty}P(|X_n-X|\geq\epsilon)=0$$

依分布收敛：随机变量序列 $\{X_n\}$ 依分布收敛于 X，记作 $X_n\xrightarrow{d}X$，如果对任意 x，$F_{X_n}(x)\to F_X(x)$。**依概率收敛可以推出依分布收敛，反之不亦然**

几乎必然收敛：随机变量序列 $\{X_n\}$ 几乎必然收敛于 X，记作 $X_n\xrightarrow{a.s.}X$，如果 $\forall\epsilon>0$

$$\lim_{n\to\infty}P\left(\bigcup_{m=n}^{\infty}|X_m-X|\geq\epsilon\right)=1$$

7 参数估计

估计量：样本的函数，用于估计未知参数。

偏差：$\mathrm{Bias}(\hat{\theta})=E(\hat{\theta})-\theta$，$\mathrm{Bias}(\hat{\theta})=0$ 称为无偏估计量。若 $\lim_{n\to\infty}E(\hat{\theta})=\theta$，称为**渐近无偏**估计量。

均方误差：$\mathrm{MSE}(\hat{\theta})=E((\hat{\theta}-\theta)^2)=\mathrm{Var}(\hat{\theta})+(\mathrm{Bias}(\hat{\theta}))^2$。若无偏则 $\mathrm{MSE}(\hat{\theta})=\mathrm{Var}(\hat{\theta})$。

一致估计量：若估计量 $\hat{\theta}_n\xrightarrow{P}\theta$，则称 $\hat{\theta}_n$ 为参数 θ 的一致估计量。等价于 $\mathrm{MSE}\to 0$。

k 阶矩：$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$，$A_k$ 是 $\mu_k=E(X^k)$ 的无偏估计量，且一致。k 阶中心矩：$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$。$B_2$ 是 σ^2 的渐近无偏估计量，且一致，但不是无偏估计量。因为 $E(B_2)=E(X^2)-E(\overline{X}^2)$，而 $E(\overline{X}^2)=(E(X))^2+\mathrm{Var}(\overline{X})$（平方的期望减期望的平方），然后 $\mathrm{Var}(\overline{X})=\mathrm{Var}(X)/n$，所以 $E(B_2)=\frac{n-1}{n}\mathrm{Var}(X)$。**样本方差** $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 是 $\mathrm{Var}(X)$ 的无偏估计量。

矩法：用样本矩替换总体矩。**MLE**：最大化似然函数 $L(\theta)=P(X_1=x_1,\cdots,X_n=x_n)$。

区间估计：设计统计量 $\hat{\theta}_L(X_1,\cdots,X_n)$ 和 $\hat{\theta}_U(X_1,\cdots,X_n)$，使得 $P(\hat{\theta}_L\leq\theta\leq\hat{\theta}_U)\geq 1-\alpha$。

方法：**枢轴量法**。设计枢轴量 G 使得 G 的分布与未知参数无关，然后选择 c,d 使得 $P(c\leq G\leq d)=1-\alpha$，从而得到不等式 $c\leq G(X_1,\cdots,X_n,\theta)\leq d$，解出 θ 的区间估计。

8 回归分析

回归分析：$y=\alpha+\beta x+\epsilon$，其中 ϵ 为误差项，$E(\epsilon)=0$，$\mathrm{Var}(\epsilon)=\sigma^2$。

最小二乘：$Q(\alpha,\beta)=\sum_{i=1}^{n}(y_i-\beta x_i-\alpha)^2$，使得 Q 最小的 $\hat{\alpha},\hat{\beta}$ 称为最小二乘估计。求偏导然后令为 0 可得

$$\hat{\beta}=\frac{S_{xy}}{S_{xx}},\hat{\alpha}=\overline{y}-\hat{\beta}\overline{x}$$

其中 $S_{xy}=\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})$，$S_{xx}=\sum_{i=1}^{n}(x_i-\overline{x})^2$。

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer nunc elit, tristique quis dignissim sit amet, malesuada nec ante. Phasellus quis tincidunt turpis. Nullam porttitor mi ut sem condimentum, eget malesuada sapien lacinia. Nunc sagittis porttitor interdum. Proin sit amet mattis tortor. Fusce vulputate quam nec vestibulum placerat. Nunc accumsan vestibulum tortor, a accumsan diam tempus vitae. Cras varius molestie purus vitae fermentum. Nullam malesuada congue luctus. Vivamus quis neque neque. Donec ornare dictum odio, quis vestibulum ligula elementum aliquam.

Cras vitae pharetra elit, sit amet finibus purus. Maecenas at mauris et turpis faucibus vestibulum. Nulla facilisi. Vivamus faucibus ullamcorper ante, quis condimentum orci imperdiet pretium. Praesent sed quam vulputate nibh elementum lobortis eu tempor ligula. Vivamus non urna elit. Morbi imperdiet magna pulvinar tellus volutpat rutrum. Mauris imperdiet purus ullamcorper, pellentesque ligula sit amet, ultricies lacus. Nulla luctus euismod egestas. Nam eleifend nunc a enim sagittis, feugiat auctor libero luctus. Nullam nec placerat sem, in aliquet tellus.

Duis non elementum neque. Vivamus eget nibh vel neque mattis tincidunt. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Maecenas et quam vitae enim tempus luctus eu sit amet enim. In facilisis ex eget vehicula vestibulum. Phasellus laoreet arcu in nibh sodales semper vitae sit amet leo. Vestibulum sagittis pellentesque quam id imperdiet. Cras tempor consequat tellus. In congue facilisis varius. Duis tempor felis eget ante dignissim, sed ultrices turpis lacinia. Nunc nec nunc nulla. Aliquam tempor mauris vel placerat vehicula. Aliquam vulputate ut sem eget interdum. Donec a faucibus est. Vestibulum volutpat enim aliquet nisi faucibus bibendum. Sed rutrum eleifend sem, nec hendrerit tortor mattis accumsan.

Suspendisse cursus felis sed nunc egestas dapibus. Nunc quis eros sem. Ut lorem elit, vestibulum quis neque sed, tincidunt pellentesque eros. Curabitur ac condimentum lacus. Suspendisse potenti. Proin commodo lorem arcu, at porttitor ligula aliquet quis. Nulla facilisi. Integer pretium quam leo, et iaculis nisi semper et. Cras vel volutpat lectus. Phasellus porttitor, lacus vel pretium accumsan, ex lacus maximus sapien, sit amet suscipit turpis ex vel quam. Aliquam egestas porta arcu nec bibendum.

Mauris facilisis tortor nec lectus accumsan mattis. Donec eget commodo elit, non porttitor neque. Cras lobortis orci quis lacus consequat, sed rhoncus augue iaculis. Donec quis tincidunt ex. Curabitur in elit tellus. Integer pharetra dui metus, nec auctor nisi gravida eget. Fusce orci nulla, laoreet tincidunt varius accumsan, laoreet at neque. Aenean at mauris venenatis metus semper eleifend.