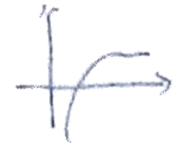
1. 数学分析

① log函数：$y=\log_e x$　$\frac{dy}{dx}=\frac{1}{x}$　$x=1$ 时 $\log_e x$ 斜率为1 

② 求极限

③ 单调递增有上界的函数 ⇒ 有极限

④ 导数：一阶导数反映曲线变化的快慢；二阶导数反映斜率变化的快慢，表征曲线的凸凹性，加速度总是向着曲线凹的一面

⑤ 常用的导数.

$C'=0$　$(x^n)'=nx^{n-1}$　$(\sin x)'=\cos x$　$(\cos x)'=-\sin x$

$(a^x)'=a^x\ln a$　$(e^x)'=e^x$　$(\log_a x)'=\frac{1}{x}\log_a e$　$(\ln x)'=\frac{1}{x}$

$(u+v)'=u'+v'$　$(uv)'=u'v+v'u$

⑥ 积分可以看作导数的逆运算：对导数求积分就是函数本身
对积分求导数也是函数本身

⑦ 分部导数公式 ⇔ 分部积分公式

$(uv)'=u'v+uv'$ ⇔ $\int f(x)\,dx = \underbrace{f(x)\cdot x}_{(u\ \ v)'} - \int x\,df(x)$

$f(x)\cdot x=\int f(x)\,dx+\int x\,df(x)$

⑧ 幂指函数 $f(x)=x^x$ 求极值.
函数先降后升，如何求最小值.

$\ln(f(x))=\ln(x^x)$

$\xrightarrow{两边求导}$ $\frac{1}{f(x)}f(x)'=\ln x+1$

$\xrightarrow{令导数为0}$ $\ln x+1=0$

$x=e^{-1}$ ⇒ 得到驻点 ⇒ 再求二阶导 说明驻点是极值

$\ln(x)=e^{-\frac{1}{e}}$

⑨ $\ln(N!)$ 数量级是 $N\ln N - N$

$$\ln(N!) = \sum_{i=1}^{N} \ln i \approx \int_1^N \ln x \, dx$$

$$= x\ln x\Big|_1^N - \int_1^N x \, d\ln x$$ // 分部积分公式 ⇒ 数分⑦

$$= N\ln N - \int_1^N x \frac{1}{x} dx$$

$$= N\ln N - x\Big|_1^N$$

$$= N\ln N - N$$

⑩ 泰勒展开式，麦克劳林展开式

泰勒. $f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0) + \cdots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + R_n(x)$

麦克劳林: $f(x) = f(0) + f'(0)(x-x_0) + \frac{f''(0)}{2!}(x-x_0) + \cdots + \frac{f^{(n)}(0)}{n!}x^n + o(x^n)$

函数在某个点 $x=x_0$ 处 n 阶可导，就可用泰勒展开式展开，在 $x=0$ 处 n 阶可导就可用麦克劳林展开式展开

⑪ 求 $\sin x$, e^x, … 的~~的极~~值. 用泰勒展开式求极值

⑫ 求 Gini 系数：用泰勒展开式求解.

$\sum_{k=1}^{K} p_k(1-p_k)$: Gini 系数其实是熵公式中的 $\ln(x)$ 在 $x=1$ 处一阶展开

熵 $H(x) = -\sum_{k=1}^{K} p_k \boxed{\ln p_k} = \sum_{k=1}^{K} p_k(-\ln p_k)$

$\approx \sum_{k=1}^{K} p_k \boxed{(1-p_k)}$ ← $f(x) = -\ln(x)$ 在 $x=1$ 处一阶展开 忽略高阶无穷小

⑬ 方向导数：$z=f(x,y)$ 在点 (x,y) 可微，则函数在该点任一方向 l 导数都存在，φ 记为 ~~[illegible]~~ x 轴到方向 l 的转角.

$$\frac{\partial f}{\partial l} = \frac{\partial f}{\partial x}\cos\varphi + \frac{\partial f}{\partial y}\sin\varphi = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \cdot \begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix}$$

⑭ 对任意方向l_2，何时在方向l上导数最大：l_2与l方向相同时，即用于梯度下降

⑮ 梯度：$z=f(x,y)$在平面区域D内具有一阶连续偏导数

对区域内每个点$P(x,y)$，向量为

向量$\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$为$z=f(x,y)$在点$P$的梯度，记做$\mathrm{grad}\, f(x,y)$

梯度方向是函数在该点变化最快的方向

⑯ Euler（欧拉Γ函数）：阶乘在实数域的推广

$$\Gamma(x) = (x-1)\,\Gamma(x-1)$$

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t}\, dt = (x-1)!$$

普通幂指函数与普通指数函数相乘

⑯ 凸函数（convex）

$\forall x, y \in \mathrm{dom}\, f$，$0 \le \theta \le 1$，有

$$f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)$$

一阶可微时：凸函数充要条件是任意一点，函数值大于等于经过该点的切线

二阶可微时：凸函数充要条件是二阶导数为正（对一元函数）或二阶梯度矩阵正定（对多元函数）

e.g. $f(x,y) = 3x^2 + 4y^2 - xy$

$$\nabla f = \begin{pmatrix} 6x - y \\ 8y - x \end{pmatrix} \qquad H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}, & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x}, & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 6, & -1 \\ -1, & 8 \end{pmatrix}$$

对x,y求偏导　　Hessian矩阵　二阶梯度矩阵

⑰ 正定：正数在N维空间的推广

一阶顺序主子式 $6>0$

二阶顺序主子式 $47>0$

正定

⑱ 凸函数举例

指数 $f(x) = e^{ax}$

幂 $f(x) = x^a$　$x \in R$　$a \ge 1$或$a \le 0$

负对数 $f(x) = -\ln x$

最大值 $f(\vec{x}) = \max(x_1, x_2, \dots, x_n)$

负熵 $f(x) = x\ln x$

范数 $f(\vec{x}) = \|x\|$

指数线性 $f(\vec{x}) = \log(e^{x_1} + e^{x_2} + \dots + e^{x_n})$

2 概率论 & 贝叶斯

① PDF

$P(X=x)$：概率，概率分布率/概率质量（离散函数），概率密度（连续函数）

$P(X=x)=0$ 不代表一定不发生，这是一个求极限的概念

② CDF

累积分布函数 $\Phi(x)=P(x\le x_0)$

对于连续函数 $\Phi(x)=P(x\le x_0)=\int_{-\infty}^{x_0}p(x)dx$

对于离散函数 $\Phi(x)=P(x\le x_0)=\sum_{x=-\infty}^{x_0}p(x)$

值域在 [0,1] 的单调递增函数，有可能是概率函数（累积分布），例如

sigmoid函数 $F(x)=\frac{1}{1+e^{-x}}$ $(x\in(-\infty,+\infty))$

③ 古典概率 $P(A)=P_N^n/N^n$ n个球放入N个箱子，每个箱子至多有一个球概率（事件A）

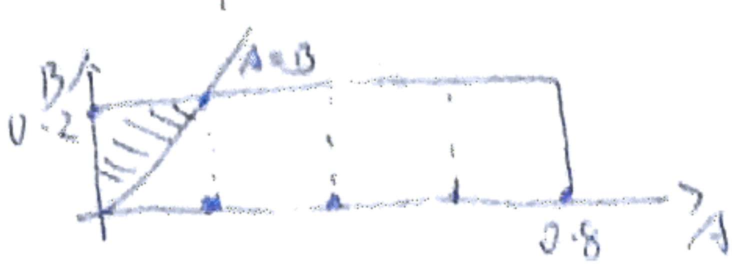
分母：从球1到球n，每个球有N种放法

分子：因为每个盒子至多放一个球，球1有N种放法，球2有N-1种，…

$P_N^n=N(N-1)(N-2)\cdots(N-n+1)$

用途1：生日悖论：50个人至少2人生日相同的概率

用途2：商品推荐，商品A（强相关）随机生成0~0.8的分数，商品B（增加新颖度）随机生成0~0.2的分数，B>A的概率



$P(B>A)=\frac{1}{1+7}=0.125$

④ 熵：$H(P)=-\sum_{k=1}^{K}(P_k\ln P_k)$ 熵越高表示混乱程度越高；值越低混乱程度越低

（H(P) 取最高值 相当于分到K个类别概率均等，即完全混乱）

H(P)是装箱问题（n个物品装K个箱子有多少种装法）让 $N\to+\infty$ 时的极限值

装箱问题：$\frac{n!}{n_1!\,n_2!\cdots n_K!}$ n个物品，$n_1,n_2,\cdots n_K$ 是箱子 $1,2,\cdots K$ 的容量

分2箱，分法是 $C_n^{n_1}=\frac{n!}{n_1!(n-n_1)!}=\frac{n!}{n_1!\,n_2!}$ 推广到分n个箱子

因为 $\lim_{N\to\infty}(\ln N!)=N(\ln N-1)$

$\frac{1}{n}\ln\left(n!/\prod n_k!\right)=-\sum_{k=1}^{K}\left[\frac{n_k}{n}\ln\frac{n_k}{N}\right]=-\sum_{k=1}^{K}(P_k\ln P_k)=H(P)$

⑤ 概率和直观有时相违背：9点概率（本福特定律），公路堵车模型

⑥ 条件概率：$P(Y|X) = P(YX)/P(X)$

全概率公式：$P(Y) = \sum_i P(Y|X_i) \cdot P(X_i)$

贝叶斯公式（Bayes）

似然函数，给定特征预测分类结果

$$\boxed{P(X|Y)} = \frac{\boxed{P(Y|X)} \cdot \boxed{P(X)}}{\boxed{P(Y)}}$$

$P(X)$ → 特征（如是否校准枪械）

$P(Y)$ → 样本分类结果（如是否命中靶心）

$P(X|Y)$：后验概率，给定分类结果时特征取值的概率。

$$P(X_i|Y) = \frac{P(X_iY)}{P(Y)} = \frac{P(Y|X_i) \cdot P(X_i)}{\sum_j P(Y|X_j) \cdot P(X_j)}$$

⑦ 两点分布（Bernoulli Distribution）：即 0-1 分布

X	1	0
Y	p	1-p

$E(X) = 1 \cdot p + 0 \cdot q = p$

$D(X) = E((X-E(X))^2) = E(X^2 - 2E(X) \cdot X + (E(X))^2)$
方差

$= E(X^2) - (E(X))^2 = 1^2 \cdot p - p^2 = p \cdot q$

⑧ 二项分布（Binomial Distribution）：参数 n, p，对服从参数 p 的二项分布样本做 n 次试验，即 $X = \sum_{i=1}^{n} X_i$ 其中 X_i 服从参数为 p 的二项分布。

$E(X) = \sum_{i=1}^{n} E(X_i) = n \cdot p$

$D(X) = \sum_{i=1}^{n} D(X_i) = n \cdot p \cdot q = n \cdot p(1-p)$
方差

$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

⇒ PPT
★ 从定义角度计算
$E(X)$，$D(X)$

⑨ 负二项分布：对一系列服从参数为 p（成功概率）的0-1分布的独立实验，持续直到 r 次成功，实验次数 X 的概率：

$$P(X=x, r, p) = C_{x-1}^{r-1} \cdot p^r (1-p)^{x-r} \quad x \in [r, r+1, \dots, \infty)$$

若记 $x-k$ 为失败次数

$$P(X=k, r, p) = C_{k+r-1}^{r-1} \cdot p^r (1-p)^k \quad k \in N$$

⑩ 泊松分布：一段时间/一个空间内某个事件出现的个数 k　$X \sim \pi(\lambda)$

泊松分布参数 数学期望 方差

$$P\{X=k\} = \frac{\lambda^k}{k!}e^{-\lambda} \quad k=0,1,2,\cdots,\lambda>0$$

$$E(X) = \sum_{k=0}^{\infty} k\cdot\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}\cdot\lambda = \lambda e^{-\lambda}e^{\lambda} = \lambda$$

$D(X) = \lambda$　//推导见 PPT

泰勒展开式：$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^k}{k!} + R_k$

$$1 = 1\boxed{e^{-x}} + \boxed{x\cdot e^{-x}} + \boxed{\frac{x^2}{2!}\cdot e^{-x}} + \boxed{\frac{x^3}{3!}e^{-x}} + \cdots \boxed{+\frac{x^k}{k!}e^{-x}} + R_n\cdot e^{-x}$$

↑离散分布
↓连续分布

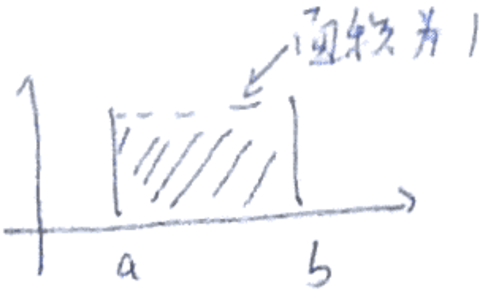
就是泊松分布的概率密度　$P\{X=1\}$　$P\{X=2\}$　$P\{X=3\}$　$P\{X=k\}$

---

⑪ 均匀分布：$X \sim U(a,b)$

$$f(x) = \begin{cases} \frac{1}{b-a} & a<x<b \\ 0 & 其它 \end{cases}$$

概率密度



$$E(X) = \int_{-\infty}^{\infty} x\cdot f(x)\,dx = \int_a^b \frac{1}{b-a}\cdot x\cdot dx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b+a}{2}$$

$$D(X) = E(X^2) - (E(X))^2 = \int_a^b x^2\frac{1}{b-a}dx - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

⑫ 指数分布 (Exponential Distribution)

$$f(x) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}} = \lambda e^{-\lambda x} & x>0 \quad 其中\theta>0 \\ 0 & x\le 0 \end{cases}$$

强度　速率

期望　$\boxed{在(-\infty,+\infty)积分为1}$

概率密度

$$E(X) = \int_{-\infty}^{+\infty} x\cdot f(x)\,dx = \int_0^{+\infty} x\cdot\frac{1}{\theta}e^{-\frac{x}{\theta}}dx = \cdots = \theta$$

$$D(X) = \cdots = \theta^2$$

无记忆性：
eg. $P\{X>s+t \mid X>s\} = P\{X>t\}$
电子元件使用t小时故障
与使用s小时后再使用t小时
故障概率相同
(不是所有的分布都
符合式子确定要求建
模成指数分布的)

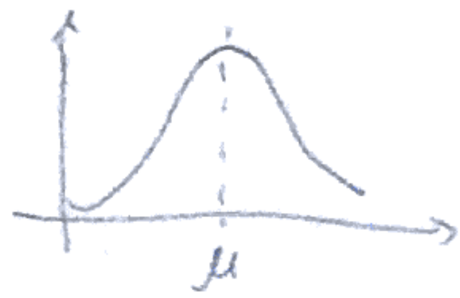
站台人数 ~ 泊松分布
人的等待时间 ~ 指数分布

⑬ 正态分布．$X\sim N(\mu,\sigma^2)$，μ表示分布中心，σ^2是方差，$\frac{1}{\sigma^2}$是精度

(在分布中心聚集程度

$$f(x)=\frac{1}{(\sqrt{2\pi})\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\qquad \sigma>0,\ -\infty<x<+\infty$$

概率密度



$$E(X)=\int_{-\infty}^{+\infty}x\cdot\frac{1}{(\sqrt{2\pi})\cdot\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

$$=\int_{-\infty}^{+\infty}(\mu+\sigma t)\cdot\frac{1}{(\sqrt{2\pi})\sigma}e^{-\frac{1}{2}t^2}d(\mu+\sigma t)$$

令 $\frac{x-\mu}{\sigma}=t$ 则 $x=\mu+\sigma\cdot t$.

二元正态分布 $\longrightarrow \begin{bmatrix}\sigma_{xx}^2 & \sigma_{xy}^2\\ \sigma_{yx}^2 & \sigma_{yy}^2\end{bmatrix}$

⑭ 两点/二项/泊松/均匀/指数/正态分布参数，数学期望，方差.

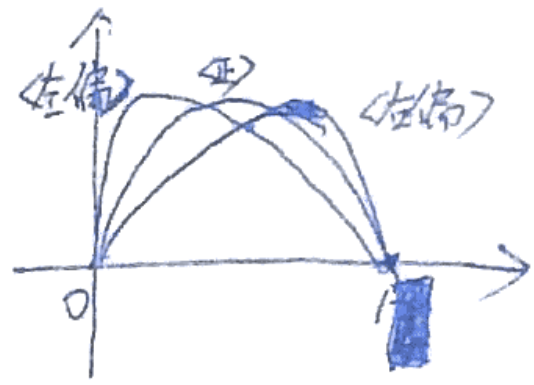
两点	$0<p<1$	$E(X)=p$	$D(X)=p(1-p)$
二项	$n\geq 1$, $0<p<1$	$E(X)=np$	$D(X)=n\cdot p\cdot(1-p)$
泊松	$\lambda>0$	$E(X)=\lambda$	$D(X)=\lambda$
均匀	$a<b$	$E(X)=\frac{1}{2}(a+b)$	$D(X)=(b-a)^2/12$.
指数	$\theta>0$	$E(X)=\theta$	$D(X)=\theta^2$
正态	μ, $\sigma>0$	$E(X)=\mu$	$D(X)=\sigma^2$.
Beta分布	α, β	$E(X)=\frac{\alpha}{\alpha+\beta}$	

⑮ Beta分布与Gamma函数

a) 抛物线 <正><左偏><右偏> $x\in[0,1]$

$\boxed{\alpha=2,2,3}$ $\boxed{\beta=2,3,2}$

$$l(x)=\begin{cases} x^{(2-1)}\cdot(1-x)^{(2-1)} & \text{〈正〉} \\ x^{(2-1)}\cdot(1-x)^{(3-1)} & \text{〈左偏〉} \\ x^{(3-1)}\cdot(1-x)^{(2-1)} & \text{〈右偏〉} \end{cases}$$



b) 抛物线下面积: $B(\alpha,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx=\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

Gamma函数: $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}dt=(x-1)!$ 阶乘在实数域推广

c) 概率密度: 抛物线y值除以抛物线下面积, 值域落在[0,1]

$$f(x)=\begin{cases}\dfrac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1} & x\in[0,1] \\ 0 & \text{其它}\end{cases}$$

d) 数学期望

$$E(x)=\int_0^1 x\cdot\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx=\boxed{\frac{\alpha}{\alpha+\beta}}$$

决定左偏/右偏程度

⑯ 指数族 (The exponential family) 函数可写成

$$f(x) = b(x) \cdot e^{\eta \cdot g(x) - \alpha(\eta)}$$

η → (natural parameter / canonical parameter)

$$\overset{\text{即}}{=} \underbrace{b(x)}_{\text{归一化}} \cdot \exp(\underbrace{\eta^T}_{\text{连接数}} \underbrace{T(x)}_{\substack{\text{连接函数} \\ \text{sufficient statistic}}} - \underbrace{\alpha(\eta)}_{\text{log partition function}})$$

$e^{-\alpha(\eta)}$: normalization constant 正则化系数.

<a> 贝努里(二项)分布写成指数族：得到 sigmod 函数.

$$\underset{\text{概率密度}}{f(x)} = p^x (1-p)^{1-x} = \exp(\ln(p^x (1-p)^{1-x}))$$

$$= \exp(x\ln(p) + (1-x)\ln(1-p)) = \exp\big((\underbrace{\ln\frac{p}{1-p}}_{\text{记为}\eta})x + \underbrace{\ln(1-p)}_{\text{记为}\alpha(\eta)}\big)$$

记为 η，则 $p = \frac{1}{1+e^{-\eta}}$

$\eta = \ln\frac{p}{1-p}$

导出 $p = \frac{1}{1+e^{-\eta}}$ // sigmod 函数

$= \frac{1}{1+e^{-\sum\theta_i x_i}}$

$$= \underbrace{1}_{} \cdot \exp(\underbrace{(\ln\frac{p}{1-p})}_{} \cdot \underbrace{x}_{} + \underbrace{\ln(1+e^{\eta})}_{})$$

$\boxed{b(x)=1}$ $\boxed{\eta = \ln\frac{p}{1-p}}$ $\boxed{T(x)=x}$ $\boxed{\alpha(\eta) = \ln(1+e^{\eta})}$

$\alpha(\eta) = \ln(1-p)$

$= \ln(1+e^{\eta})$

⇓

把 $\eta = \ln\frac{p}{1-p}$ 建模成 $\sum_{i=1}^{n}\theta_i x_i$ 即特征的线性和，就是 LR

<b> 高斯分布 (Gaussian) 也能写成指数族

$$\underset{\text{概率密度}}{f(x)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2} = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(x-\mu)^2)$$

$$= \underbrace{\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)}_{} \cdot \exp(\underbrace{\mu x}_{} - \underbrace{\frac{1}{2}\mu^2}_{})$$

$\boxed{b(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})}$ 归一化

$\boxed{\eta = \mu}$ 连接数 $\boxed{T(x)=x}$ 连接函数

$\boxed{\alpha(\eta) = \frac{-\mu^2}{2} = \frac{-\eta^2}{2}}$

贝努里分布被建模为 LR

高斯分布被建模为线性回归

⑰ 事件独立：$P(AB) = P(A)\cdot P(B)$　　$P(A|B) = P(A)$

$I(A,B) = 0$

互信息：A，B 相互包含的信息量

⑱ 数学期望：$E(X) = \sum_i x_i p_i$　/*离散事件*/

$E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx$　/*连续概率*/

$E(kX) = kE(X)$

$E(X+Y) = E(X) + E(Y)$

$E(XY) = E(X)E(Y)$　/*当且仅当 X，Y 相互独立*/

⑲ 方差 $Var(X) = E\{[X-E(X)]^2\} = E(X^2) - E^2(X)$

$Var(c) = 0$

$Var(X+c) = Var(X)$

$Var(kX) = k^2 Var(X)$

$Var(X+Y) = Var(X) + Var(Y)$

⑳ 协方差 $\boxed{Cov(X,Y) = E\{[X-E(X)][Y-E(Y)]\} = E(XY) - E(X)E(Y)}$

$Cov(X,Y) = Cov(Y,X)$

$Cov(aX+b, cY+d) = a\cdot c\, Cov(X,Y)$

$Cov(X_1+X_2, Y) = Cov(X_1,Y) + Cov(X_2, Y)$

是两个随机变量 (X, Y) 具有相同方向变化趋势的度量

$Cov(X,Y) > 0$ 变化趋势相同（X 增大，Y 也增大）

$Cov(X,Y) = 0$ 不相关

$Cov(X,Y) < 0$ 变化趋势相反

协方差上界

$|Cov(X,Y)|^2 \le Var(X)\cdot Var(Y)$　当且仅当 $X = kY + b$ 时取等号

线性相关

$Cov^2(X,Y) = (E((X-E(X))(Y-E(Y))))^2$　协方差定义

$\le E((X-E(X))^2(Y-E(Y))^2)$　方差性质

$\le E((X-E(X))^2)E((Y-E(Y))^2)$　期望性质

$= Var(X)\,Var(Y)$

理解

证明过程 见PPT

(21) 皮尔逊相关系数 (Pearson)

X与Y不相关，说明X，Y没有线性关系，但不保证没有其它数学关系

(协方差，方差层面的相关度)

因此只能称为二阶相互，不保证X，Y相互独立

但如果X，Y是二维正态随机变量(已固定为正态分布)：不相关 $\Leftrightarrow$ 相互独立

这类在协方差，方差层面的相关度

用皮尔逊相关系数来度量 (其实就是归一化之后的协方差)

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} \leq 1$$

根据协方差上界定理

对n个随机变量 $(X_1, X_2, \cdots, X_n)$

(22) 协方差矩阵：用任意两个元素 X_i, X_j 的协方差

$$C = \begin{bmatrix} Cov(X_1,X_1), & Cov(X_1,X_2), & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1), & Cov(X_2,X_2), & \cdots & Cov(X_2,X_n) \\ \vdots & \vdots & & \vdots \\ Cov(X_n,X_1), & Cov(X_n,X_2), & \cdots & Cov(X_n,X_n) \end{bmatrix}$$

(23) 贝努里版本的大数定理：事件A发生的~~概率~~频率 n_A/n 收敛于该事件概率P

用途：(a) 正态分布参数估计

(b) 朴素贝叶斯做垃圾邮件分类

(c) 隐马尔可夫模型有监督参数学习

证明过程：切比雪夫不等式 $\Rightarrow$ 大数定理 $\Rightarrow$ 大数定理推论 $\Rightarrow$ 贝努里定理

(24) 最大似然估计：样本 $X_1, X_2, \cdots X_n$ 服从参数未知的概率分布

$\theta_1, \theta_2 \cdots \theta_k$ (高斯分布的均值，方差 / 泊松分布的λ…)

联合概率密度

$$\underbrace{L(x_1, x_2, \cdots x_n}_{\text{固定}}, \theta_1, \theta_2 \cdots \theta_k) = \prod_{i=1}^{n} f(\underbrace{x_i}_{\text{固定}}, \underbrace{\theta_1, \theta_2, \cdots \theta_k}_{\text{未知}})$$

找到让这个联合概率密度L取最大值的参数 $\theta_1, \theta_2, \cdots \theta_k$

就是最大似然估计

㉕ 最大似然估计求解：求函数 L 最大值，需求导，$\prod_{i=1}^{n} f(x_i, \theta_1, \dots \theta_k)$ 乘积不方便求导 ⇒ 对 log(L) 对数似然函数求最值

$$\log L(\theta_1, \theta_2, \dots \theta_k) = \sum_{i=1}^{n} \log f(x_i, \theta_1, \dots \theta_k)$$

$$\frac{\partial L(\theta)}{\partial \theta_i} = 0, \quad i = 1, 2, \dots k$$

求偏导，让偏导为0得到驻点，分析该驻点，因为大点

㉖ 例子：对高斯分布做最大似然估计

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

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

$$l(x) = \log(L(x)) = \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

目标函数：对数似然函数

$$= \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi}\sigma} + \sum_{i=1}^{n} \log\left(e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right)$$

$$= \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi}\sigma} + \sum_{i=1}^{n} -\frac{(x_i-\mu)^2}{2\sigma^2}$$

$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i (x_i - \mu)^2$$

对 $l(x)$ 关于 σ^2, μ 求偏导，得到

$\mu = \frac{1}{n}\sum_i x_i$ /* 样本均值 */

$\sigma^2 = \frac{1}{n}\sum_i (x_i - \mu)^2$ /* 样本的伪方差 */

结论：样本的均值即高斯分布的期望
样本的伪方差即高斯分布的方差

↓

用于 EM 算法 ，高斯混合模型.

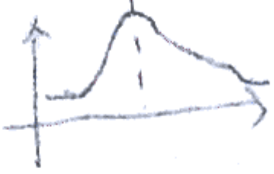
㉗ 过拟合惩罚参数 $P_男 = \frac{N_男 + 5}{N_男 + N_女 + 10}$ 更多内容见拉普拉斯平滑

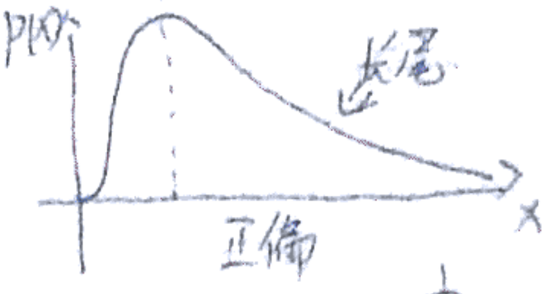
$P_女 = \frac{N_女 + 5}{N_男 + N_女 + 10}$

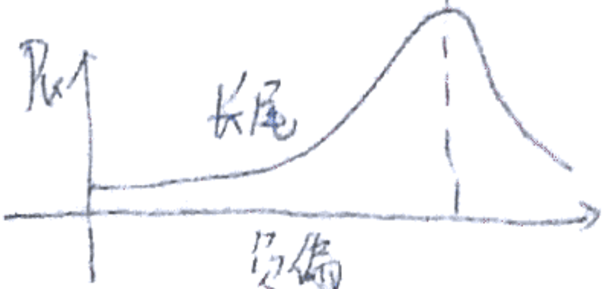
㉘ 偏度，峰度，Γ函数，熵率

※偏度：度量随机变量（相对于均值）概率分布的不对称性

偏度为
- 正：表示正偏，长尾在右侧
- 零：相对均匀地分布在均值的两侧（不保证对称）
- 负：表示负偏，长尾在左侧







偏度公式：三阶累积量与二阶累积量1.5次方的比率

$$\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{E[(X-\mu)^3]}{E[(X-\mu)^2]^{3/2}} = \frac{\kappa_3}{\kappa_2^{3/2}}$$

用Skew[X]表示偏度

$$\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{E[X^3]-3\mu E[X^2]+2\mu^3}{\sigma^3} = \frac{E[X^3]-3\mu\sigma^2-\mu^3}{\sigma^3}$$

※峰度：度量概率密度在均值处峰值高低

峰度公式：4阶中心矩除以方差的平方减3（减3为了让正态分布峰度为0）

$$\gamma_2 = \frac{\kappa_4}{\kappa_2^2} = \boxed{\frac{\mu_4}{\sigma_4}} - 3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2} - 3$$

也被称为超值峰度（excess kurtosis）

超值峰度为
- 正：尖峰态
- 零：
- 负：低峰态

※标准正态分布的超值峰度为3，峰度为0，偏度为0，参考PPT

※Γ函数：$\Gamma(x) = (x-1)\Gamma(x-1) \Rightarrow \frac{\Gamma(x)}{\Gamma(x-1)} = x-1$ 阶乘在实数域的推广

※Pearson相关系数：$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$，X与Y线性相关性等等

※FFT变换

※卷积及应用示例