初探新高考概率统计学走势与前沿瞭望

发言人：镇江 朱世耀（网名朱继光）

马克思主义哲学告诉我们世界的发展既有确定的必然性，也有不确定的偶然性，这种确定的必然性使得人类发展有条不紊，同时不确定的偶然性又使得缔造历史的过程变得扑朔迷离，扣人心弦，令人捉摸不透。概率统计学似乎正是为了描绘这种不可捉摸的美才应运而生。它需要通过严肃的数量的来述说“究竟多少才算一堆?”，“功过是否真的相抵？”，“黑与白之间到底误差多少？”这类穷尽一生也未必讲得清的哲学问题！从这角度来看的话，学习概率统计学就是学习一种科学的世界观也是恰如其分的。

个人在进修学习的过程中，深感浅薄，跟着前人的脚印侥幸获得了一点点心得，面对新高考改革不可阻挡的大潮流，借助今天这次机会，本着共同学习，共同进步的理念，发表一点浅陋的见解。我将从常规的主要考点出发分别选取几个实例给大家展示，并适当的延伸。由于时间仓促，定有疏漏和不足之处，欢迎大家及时指正批评。

1．事件基本概念：

$$\begin{cases}\text{互斥事件} \Leftrightarrow A\cap B=\varnothing \Leftrightarrow \text{事件}A\text{ 和}B\text{ 不能同时发生}\\ \text{对立事件} \Leftrightarrow A\cap B=\varnothing\text{且}A\cup B=S \Leftrightarrow A,B\text{ 中有且仅有一个发生}\end{cases}$$

口诀：互斥未必互逆，互逆一定互斥，互斥必不独立.

2．逻辑集合论：

$$\begin{cases}\neg(p\wedge q)=\neg p\vee\neg q\\ \neg(p\vee q)=\neg p\wedge\neg q\\ \complement_U(A\cap B)=\complement_U A\cup\complement_U B\\ \complement_U(A\cup B)=\complement_U A\cap\complement_U B\end{cases}$$

3．概率的加法运算公式

$$P(A\cup B)=P(A)+P(B)-P(AB)$$

4．放回抽样与不放回抽样的联系

（参考资料）袋中有a只白球，b只红球，k个人依次在袋中取一只球，(1) 作放回抽样 (2) 作不放回抽样，求第i（$i=1$，2，…，k）人取到白球（记为事件B）的概率（$k\leqslant a+b$）.

解：(1) 放回抽样的情况，显然有 $P(B)=\dfrac{a}{a+b}$，每次取球均为等可能独立重复事件.

（2）不放回抽样的情况．各人取一只球，每种取法是一个基本事件，共有 $(a+b)(a+b-1)\cdots(a+b-k+1)=A_{a+b}^{k}$个基本事件，且由于对称性每个基本事件发生的可能性相同. 当事件B发生时，第i人取的是白球，它可以是a只白球中的任一只，有a种取法. 其余被取的$k-1$只球可以是其余$a+b-1$只球中的任意$k-1$只，共有 $(a+b-1)(a+b-2)\cdots[a+b-1-(k-1)+1]=A_{a+b-1}^{k-1}$ 种取法，于是事件B中包含a $\cdot A_{a+b-1}^{k-1}\Big/A_{a+b}^{k}=\dfrac{a}{a+b}$.

值得注意的是$P(B)$与i无关，即k个人取球，尽管取球的先后次序不同，各人取到的白球的概率是一样的，大家机会相同（例如在买彩票时，各人得奖机会相同）. 另外放回抽样与不放回抽样情况下$P(B)$是一样的．

5．搭配组排问题

（参考资料）将 15 名新生随机的平均分配到三个班级中，这 15 名新生中有3名是优秀学生．问（1）每个班级各分配到一名优秀生的概率是多少？（2）3 名优秀生分配在同一个班级的概率是多少？

解： 15 名新生分配到三个班级的分法总数： $N=C_{15}^{5}C_{10}^{5}C_{5}^{5}$

（1）3 名优等生平均分到三个班时，分法总数：$N_1=A_3^3C_{12}^4C_8^4C_4^4$

故所求概率为： $p_1=\dfrac{N_1}{N}=\dfrac{25}{91}$；

（2）3 名优等生去全在一个班时，分法总数：$N_2=A_3^1C_{12}^2C_{10}^5C_5^5$

故所求概率为： $p_2=\dfrac{N_2}{N}=\dfrac{6}{91}$．

例 1．（2010 江西·理）6 为志愿者分成 4 组，其中两个组各 2 人，另两个组各 1 人，分赴世博会四个不同场馆服务，不同的分配方案有多少种？

解：可以分组的情况仅有 2，2，1，1 这一种，共有 $\frac{C_6^2C_4^2C_2^1C_1^1}{A_2^2A_2^2}$ 种分法（注意去掉同数组组之间的差别），再考虑到将分好的 4 组，任意分配到 4 个不同的场馆，即完成一次全排列 A_4^4．因此总的分配方案为 $\frac{C_6^2C_4^2C_2^1C_1^1}{A_2^2A_2^2}A_4^4$ 种．

6．条件概率（预测未来改革考点，在模考中会大面积出现，真题中也会顺带考察）

全概率公式与贝叶斯公式的简单综合运用．

公式：$P(B|A)=\frac{P(AB)}{P(A)}=\frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|\bar{B})P(\bar{B})}$．

（参考资料）据资料报道，在美国患肺癌的概率约为 0.1%，在人群中有 20% 是吸烟者，他们患肺炎的概率约为 0．4％，求不吸烟者患肺癌的概率是多少？

解：以 C 记事件“患肺癌”，以 A 记事件“吸烟”，按题意 $P(C)=0.001$，$P(A)=0.20$，

$P(C|A)=0.004$．需要求条件概率 $P(C|\bar{A})$．由全概率公式有

$P(C)=P(C|A)P(A)+P(C|\bar{A})P(\bar{A})$．

代入数据得：0．$001=0.004\times0.20+P(C|\bar{A})P(\bar{A})=0.004\times0.20+P(C|\bar{A})\times0.80$，

$P(C|\bar{A})=0.00025$．

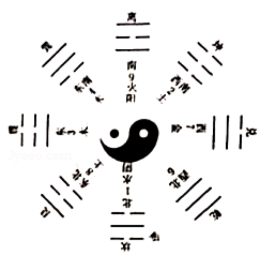
例 2．（20 年辽宁三模）2020 年初，新型冠状肺炎在欧洲爆发后，我国第一时间内向相关国家捐助医疗物资，并派出由医疗专家组成的医疗小组奔赴相关国家．现有四个医疗小组甲、乙、丙、丁，和有 4 个需要援助的国家可供选择，每个医疗小组只去一个国家，设事件 $A=$ “4 个医疗小组去的国家各不相同”，事件 $B=$ “小组甲独自去一个国家”，则 $P(A|B)$ 为多少？

解：事件 $A=$ “4 个医疗小组去的国家各不相同”，事件 $B=$ “小组甲独自去一个国家”，

则 $P(AB)=\dfrac{A_4^4}{4^4}=\dfrac{3}{32}$，$P$（B）$=\dfrac{C_4^1\cdot 3^3}{4^4}=\dfrac{27}{64}$，

$P(A|B)=\dfrac{P(AB)}{P(B)}=\dfrac{2}{9}$．

例 3．（20 年大连二模）《易经》是中国传统文化中的精髓，如图是易经后天八卦图（含乾、坤、巽、震、坎、离、艮、兑八卦），每一卦由三根线组成(▬▬ 表示一根阳线，▬ ▬ 表示一根阴线)，从八卦中任取两卦，记事件 $A=$ “两卦的六根线中恰有两根阳线”，$B=$ “有一卦恰有一根阳线”，则 $P(A|B)=$________．



后天八卦图

解：观察八卦图可知，含 3 根阴线的共有 1 卦，含 3 根阳线的共有 1 卦，还有 2 根阴线 1 根阳线的共有 3 卦，含有 1 根阴线 2 根阳线的共有 3 卦，

$\therefore$ 从八卦中任取两卦，有一卦恰有一根阳线的取法有：$C_3^1\cdot C_5^1+C_3^2=18$；

再此条件下：两卦的六根线恰有两根阳线的取法有：$C_3^2=3$ 种；

故 $P(A|B)=\dfrac{3}{18}=\dfrac{1}{6}$．

7．实际推断原理与假设检验（含最小二乘与显著性检验）

定义：概率很小的事情在一次实验中实际上几乎是不发生的．现在它在一次实验中发生了，因此有理由怀疑假设的正确性．

举例：某接待站一周来访 12 次，均发生在周二或周四。是否可以认为接待时间有规定？

假设没有规定，发生在周二周四的概率为 $\dfrac{2}{7}$，则连续发生 12 次的概率为 $(\dfrac{2}{7})^{12}=0.0000003$．这个概率几乎为 0，所以我们可以推断出假设错误。

例 4．（2019 全国Ⅰ理）为治疗某种疾病，研制了甲、乙两种新药，希望知道哪种新药更有效，为此进行动物试验．试验方案如下：每一轮选取两只白鼠对药效进行对比试验．对于两只白鼠，随机选一只施以甲药，另一只施以乙药．一轮的治疗结果得出后，再安排下一轮试验．当其中一种药治愈的白鼠比另一种药治愈的白鼠多 4 只时，就停止试验，并认为治愈只数多的药更有效．为了方便描述问题，约定：对于每轮试验，若施以甲药的白鼠治愈且施以乙药的白鼠未治愈则甲药得 1 分，乙药得 -1 分；若施以乙药的白鼠治愈且施以甲药的白鼠未治愈则乙药得 1 分，甲药得 -1 分；若都治愈或都未治愈则两种药均得 0 分．甲、乙两种药的治愈率分别记为 α 和 β，一轮试验中甲药的得分记为 X．

（1）求 X 的分布列；

（2）若甲药、乙药在试验开始时都赋予 4 分，$p_i(i=0$，1，…，8) 表示“甲药的累计得分为 i 时，最终认为甲药比乙药更有效”的概率，则 $p_0=0$，$p_8=1$，$p_i=ap_{i-1}+bp_i+cp_{i+1}(i=1$，2，…，7)，其中 $a=P(X=-1)$，$b=P(X=0)$，$c=P(X=1)$．假设 $\alpha=0.5$，$\beta=0.8$．

(i) 证明：$\{p_{i+1}-p_i\}(i=0$，1，2，…，7) 为等比数列；

(ii) 求 p_4，并根据 p_4 的值解释这种试验方案的合理性．

（1）解：X 的所有可能取值为 -1，0，1．

$P(X=-1)=(1-\alpha)\beta$，$P(X=0)=\alpha\beta+(1-\alpha)(1-\beta)$，$P(X=1)=\alpha(1-\beta)$，

$\therefore X$ 的分布列为：

X	-1	0	1
P	$(1-\alpha)\beta$	$\alpha\beta+(1-\alpha)(1-\beta)$	$\alpha(1-\beta)$

（2）(i) 证明：$\because \alpha=0.5$，$\beta=0.8$，

$\therefore$ 由（1）得，$a=0.4$，$b=0.5$，$c=0.1$．

因此 $p_i=0.4p_{i-1}+0.5p_i+0.1p_{i+1}(i=1$，2，…，7)，

故 $0.1(p_{i+1}-p_i)=0.4(p_i-p_{i-1})$，即 $(p_{i+1}-p_i)=4(p_i-p_{i-1})$，

又 $\because p_1-p_0=p_1\neq 0$，$\therefore \{p_{i+1}-p_i\}(i=0$，1，2，…，7) 为公比为 4，首项为 p_1 的等比数列；

(ii) 解：由 (i) 可得，

$$p_8=(p_8-p_7)+(p_7-p_6)+\ldots+(p_1-p_0)+p_0=\frac{p_1(1-4^8)}{1-4}=\frac{4^8-1}{3}p_1,$$

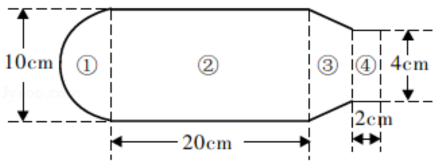
$\because p_8=1$， $\therefore p_1=\dfrac{3}{4^8-1}$，

$\therefore P_4=(p_4-p_3)+(p_3-p_2)+(p_2-p_1)+(p_1-p_0)+p_0=\dfrac{4^4-1}{3}p_1=\dfrac{1}{257}$．

P_4表示最终认为甲药更有效的概率．

由计算结果可以看出，在甲药治愈率为0．5，乙药治愈率为0．8时，认为甲药更有效的概率为$P_4=\dfrac{1}{257}\approx 0.0039$，此时得出错误结论的概率非常小，说明这种试验方案合理．

例5．（2020•福建二模）某同学使用某品牌暖水瓶，其内胆规格如图所示．若水瓶内胆壁厚不计，且内胆如图分为①②③④四个部分，它们分别为一个半球、一个大圆柱、一个圆台和一个小圆柱体若其中圆台部分的体积为$52\pi cm^3$，且水瓶灌满水后盖上瓶塞时水溢出$\dfrac{10\pi}{3}cm^3$．记盖上瓶塞后，水瓶的最大盛水量为V，



（1）求V；

（2）该同学发现：该品牌暖水瓶盛不同体积的热水时，保温效果不同．为了研究保温效果最好时暖水瓶的盛水体积，做以下实验：把盛有最大盛水量V的水的暖水瓶倒出不同体积的水，并记录水瓶内不同体积水在不同时刻的水温，发现水温y（单位：$^{\circ}$C) 与时刻t满足线性回归方程$y=ct+d$，通过计算得到如表：

倒出体积 xcm^3	0	30	60	90	120
拟合结果	$y=c_1t+d$	$y=c_2t+d$	$y=c_3t+d$	$y=c_4t+d$	$y=c_5t+d$
倒出体积 xcm^3	150	180	210	$\cdots$	450
拟合结果	$y=c_6t+d$	$y=c_7t+d$	$y=c_8t+d$	$\cdots$	$y=c_{16}t+d$

注：表中倒出体积x（单位：cm^3) 是指从最大盛水量中倒出的那部分水的体积．其中：

C_1	C_2	C_3	C_4	C_5	C_6	C_7
-1.4	-1.3	-1.2	-1	-1.1	-0.9	-0.8

令$w=|c|$，$|w_i=c_i|$， $x_i=30(i-1)$， $i=1$，2，…，16．对于数据$(x_i$， $w_i)(i=1$，2，…，

7)，可求得回归直线为 $L_1: w=\beta x+\alpha$，对于数据 $(x_i, w_i)(i=8, 9, \ldots, 16)$，可求得回归直线为 $L_2: w=0.0009x+0.7$.

(i) 指出 $|c|$ 的实际意义，并求出回归直线 L_1 的方程（参考数据：$\frac{9}{2800}\approx 0.0032$；）

（ⅱ）若 L_1 与 L_2 的交点横坐标即为最佳倒出体积，请问保温瓶约盛多少体积水时（盛水体积保留整数，且 π 取3.14) 保温效果最佳？

附：对于一组数据 (u_1, v_1)，(u_2, v_2)，…，(u_n, v_n)，其回归直线 $v=\beta u+\alpha$ 中的斜率和截距的最小二乘估计分别为 $\beta=\frac{\sum_{i=1}^{n}(u_i-\overline{u})(v_i-\overline{v})}{\sum_{i=1}^{n}(u_i-\overline{u})^2}$，$\hat{\alpha}=\overline{v}-\hat{\beta}\bullet\overline{u}$.

解：（1）依题意得，半球的半径为 $r=5cm$，体积为 $V_1=\frac{1}{2}\times\frac{4}{3}\times 125\pi=\frac{250}{3}\pi cm^3$，

大圆柱体积 $V_2=25\pi\times 20=500\pi cm^2$，小圆柱体积 $V_3=4\pi\times 2=8\pi cm^3$，

$\therefore$ 盖上瓶塞后，水瓶的最大盛水量为 $\frac{250}{3}\pi+500\pi+8\pi+52\pi-\frac{10}{3}\pi=640\pi cm^3$.

（2）$(i)|c|$ 的实际意义为倒出 xcm^3 体积水时，暖水瓶内水的降温速率 $|c|$ 越小，

降温速率越小，保温效果越好，$|c|$ 越大，降温速率越大，保温效果越差，

$\because x_i=30(i-1)$，$i=1$，2，…，7，对于回归直线 $L_1:\omega=\beta x+\alpha$，

$\because \overline{x}=\frac{x_1+x_2+\ldots+x_7}{7}=90$，$\overline{\omega}=\frac{\omega_1+\omega_2+\ldots+\omega_7}{7}=1.1$，

$\sum_{i=1}^{7}(x_i-\overline{x})(\omega_i-\overline{\omega})=-81$，$\sum_{i=1}^{7}(x_i-\overline{x})=25200$，$\therefore \hat{\beta}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(\omega_i-\overline{\omega})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}=-\frac{81}{25200}\approx -0.0032$，

$\hat{\alpha}=\overline{\omega}-\hat{\beta}\bullet\overline{x}=1.1+0.0032\times 90=1.388$. $\therefore$ 回归直线 L_1 的方程为 $\omega=-0.0032x+1.388$.

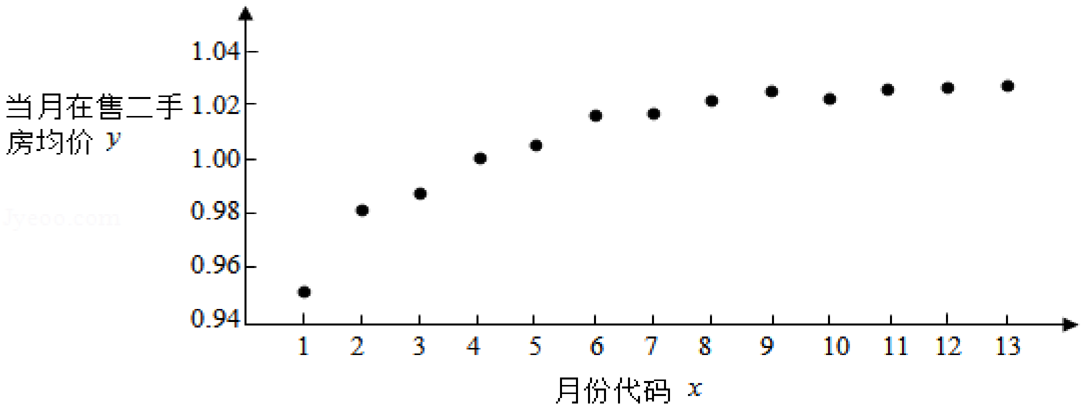
(ii) 联立 $\begin{cases}\omega=-0.0032x+1.388\\ \omega=0.0009x+0.7\end{cases}$，得 $x\approx 167.8$，

$\therefore$ 保温瓶最佳倒出体积约为 $167.8cm^3$.

保温瓶盛水体积约为 $640\pi-167.8\approx 640\times 3.14-167.8=1841.8cm^3$，

$\therefore$ 保温瓶盛水体积约为 $1841.8cm^3$ 时保温效果最佳.

例 6．（2019•河南模拟）如图是某小区 2017 年 1 月至 2018 年 1 月当月在售二手房均价（单位：万元/平方米）的散点图．（图中月份代码 1－13 分别对应 2017 年 1 月－2018 年 1 月）



根据散点图选择 $y=a+b\sqrt{x}$ 和 $y=c+dlnx$ 两个模型进行拟合，经过数据处理得到两个回归方程分别为 $\hat{y}=0.9369+0.0285\sqrt{x}$ 和 $\hat{y}=0.9554+0.0306lnx$，并得到以下一些统计量的值：

	$\hat{y}=0.9369+0.0285\sqrt{x}$	$\hat{y}=0.9554+0.0306lnx$
残差平方和 $\sum_{i=1}^{13}(y_i-\hat{y}_i)^2$	0．000591	0．000164
总偏差平方和 $\sum_{i=1}^{13}(y_i-\bar{y})^2$	0．006050	

（1）请利用相关指数 R^2 判断哪个模型的拟合效果更好；

（2）某位购房者拟于 2018 年 6 月份购买这个小区 $m(70\leqslant m\leqslant 160)$ 平方米的二手房（欲购房为其家庭首套房）．若购房时该小区所有住房的房产证均已满 2 年但未满 5 年，请你利用（1）中拟合效果更好的模型解决以下问题：

(i) 估算该购房者应支付的购房金额．（购房金额＝房款＋税费；房屋均价精确到 0．001 万元/平方米）

(ii) 若该购房者拟用不超过 100 万元的资金购买该小区一套二手房，试估算其可购买的最大面积．（精确到 1 平方米）

附注：根据有关规定，二手房交易需要缴纳若干项税费，税费是按房屋的计税价格进行征收．（计税价格＝房款）

征收方式见下表：

契税 （买方缴纳）	首套面积 90 平方米以内（含 90 平方米）为1%；首套面积 90 平方米以上且 144 平方米以内（含 144 平方米）为1.5%；面积 144 平方米以上

	或非首套为3%
增值税 （卖方缴纳）	房产证未满 2 年或满 2 年且面积在 144 平方米以上（不含 144 平方米）为5.6%；其他情况免征
个人所得税 （卖方缴纳）	首套面积 144 平方米以内（含 144 平方米）为1%；面积 144 平方米以上或非首套均为1.5%；房产证满 5 年且是家庭唯一住房的免征

参考数据：$ln2\approx0.69$，$ln3\approx1.10$，$ln17\approx2.83$，$ln19\approx2.94$，$\sqrt{2}\approx1.41$，$\sqrt{3}\approx1.73$，$\sqrt{17}\approx4.12$，$\sqrt{19}\approx4.36$．参考公式：相关指数 $R^2=1-\dfrac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}$．

解：（1）模型一中，$\hat{y}=0.9369+0.0285\sqrt{x}$ 的残差平方和为 0．000591，相关指数为 $1-\dfrac{0.000591}{0.006050}\approx0.923$；模型二中，$\hat{y}=0.9554+0.0306lnx$ 的残差平方和为 0．000164，相关指数为 $1-\dfrac{0.000164}{0.006050}\approx0.973$；$\therefore$ 相关指数较大的模型二拟合效果好些；

（2）通过散点图确定 2018 年 6 月对应的 $x=18$，

代入（1）中拟合效果更好的模型二，代入计算 $\hat{y}=0.9554+0.0306lnx=0.9554+0.0306ln18$

$=0.9554+0.0306\times(ln2+2ln3)=0.9554+0.0306\times(0.69+2\times1.10)\approx1.044$（万元/平方米）；

则 2018 年 6 月份二手房均价的预测值为 1．044（万元/平方米）；

(i) 设该购房者应支付的购房金额误 h 万元，因为税费中买方只需缴纳契税，

①当 $70\leqslant m\leqslant90$ 时，契税为计税价格的 1%，故 $h=m\times1.044\times(1\%+1)=1.05444m$；

②当 $90<m\leqslant144$ 时，契税为计税价格的 1.5%，故 $h=m\times1.044\times(1.5\%+1)=1.05966m$；

③当 $144<m\leqslant160$ 时，契税为计税价格的 3%，故 $h=m\times1.044\times(3\%+1)=1.07532m$；

$\therefore h=\begin{cases}1.05444m,70\leqslant m\leqslant90\\1.05966m,90<m\leqslant144\\1.07532m,144<m\leqslant160\end{cases}$；$\therefore$ 当 $70\leqslant m\leqslant90$ 时购房金额为 $1.05444m$ 万元，当 $90<m\leqslant144$ 时购房金额为 $1.05966m$ 万元，当 $144<m\leqslant160$ 时购房金额为 $1.07532m$ 万元；

(ii) 设该购房者可购买该小区二手房的最大面积为 t 平方米，

由 (i) 知，当 $70\leqslant m\leqslant90$ 时，应支付的购房金额为 $1.05444t$，又 $1.05444t\leqslant1.05444\times90<100$；

又因为房屋均价约为 1．044 万元/平方米，所以 $t<100$，所以 $90\leqslant t<100$，由 $1.05966t\leqslant100$，

解得 $t\leqslant\dfrac{100}{1.05966}$，且 $\dfrac{100}{1.05966}\approx94.4$，所以购房者可购买该二手房的最大面积为 94 平方米．