公式速查

几种常见分布极其数字特征

名称	符号	公式$P(x=k)$/$f(x)$	期望	方差
0-1分布	$B(1,p)$	$p^k(1-p)^{1-k}$	p	$p(1-p)$
二项分布	$B(n,p)$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
泊松分布	$P(\lambda)$	$e^{-\lambda}\frac{\lambda^k}{k!}$	λ	λ
几何分布	$G(p)$	$(1-p)^k p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
均匀分布	$U(a,b)$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布	$E(\lambda)$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
标准正态分布	$N(0,1)$	$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	0	1
一般正态分布	$N(\mu,\sigma)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
自由度为 n 的卡方分布	$\chi^2(n)$	\	n	$2n$

二项分布与泊松分布的渐进

$X\sim B(n,p)\dot{\sim}P(np)$

指数分布的性质

$$X\sim E(\lambda)\land x>0\to P(X>x)=\int_x^{+\infty}\lambda e^{-\lambda x}\mathrm{d}x=e^{-\lambda x}$$

无记忆性：$a<b\to P(X>b\mid X>a)=P(x>b-a)$

一般正态分布查表

$X\sim N(\mu,\sigma^2)\to F_X(x)=\Phi(\frac{x-\mu}{\sigma})$

切比雪夫不等式

设随机变量 X 的数学期望 $E(x)=\mu$ 和方差 $D(X)=\sigma^2$ 都存在，则对于任意的正数 $\varepsilon>0$ ，有：

$$P(|X-\mu|\geq\varepsilon)\leq\frac{\sigma^2}{\varepsilon^2}$$

推论：$P(|X-\mu|<\varepsilon)\geq 1-\dfrac{\sigma^2}{\varepsilon^2}$

上 α 分位点的性质

$u_{1-\alpha} = -u_\alpha$

$t_{1-\alpha}(n) = -t_\alpha(n)$

$F_\alpha(m,n) = \frac{1}{F_{1-\alpha}(n,m)}$

正态总体抽样分布的性质

$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$

$\bar{X}-\bar{Y} \sim N(\mu_1-\mu_2, \frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n})$

$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(m-1,n-1)$

随机事件及其概率

随机试验特点

1. 可重复性：试验可以在相同条件下重复进行。
2. 可确定性：每一次试验，可能出现各种不同结果，但所有可能出现的结果事先是明确的。
3. 不确定性：每一次试验，实际只出现一种结果，至于实际出现哪一种结果，试验之前是无法预先知道的。

随机事件的关系与运算

1. 事件的包含 $A \subset B$
2. 事件的互斥 $AB = \varnothing$
3. 事件的对立 $\bar{A}$
4. 事件的和 $A \cup B$
5. 事件的差 $A - B$
6. 事件的积 $A \cap B$ （AB）

完备事件组：对于一个样本空间 Ω，若有随机事件 $A_1, A_2, \cdots, A_n$ 满足 $\bigcup\limits_{i=1}^{n} A_i = \Omega \wedge (\forall i \neq j \rightarrow A_iA_j = \varnothing)$ 则称为对样本空间的一个分割，也叫完备事件组

1. 交换律： $AB = BA, A \cup B = B \cup A$
2. 结合律： $(AB)C = A(BC), (A \cup B) \cup C = A \cup (B \cup C)$
3. 分配律： $(A \cup B)C = AC \cup BC, (AB) \cup C = (A \cup C)(B \cup C)$
4. 对偶律： $\overline{A \cup B} = \bar{A}\bar{B}, \overline{AB} = \bar{A} \cup \bar{B}$

> 124 可推广至 n 元形式

概率的公理化定义

设随机试验 E 的样本空间为 Ω ，对试验 E 的任一随机事件 A，定义实值函数 $P(A)$ ，若它满足以下三个公理：

1. 非负有界性： $0 \leq P(A) \leq 1$
2. 规范性： $P(\Omega) = 1$
3. 可列可加性： 对无穷多个两两互不相容的随机事件 $A_1, A_2, \cdots, A_n, \cdots$ 有 $P(\bigcup\limits_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} P(A_i)$

性质：

1. 不可能事件概率为 0 ，即 $P(\varnothing) = 0$
2. 有限可加性： 对有限多个两两互不相容的随机事件 $A_1, A_2, \cdots, A_n$ 有 $P(\bigcup\limits_{i=1}^{n} A_i) = \sum\limits_{i=1}^{n} P(A_i)$
3. 对任一事件 $P(A) = 1 - P(\bar{A})$
4. 对任意两个事件 A, B，若 $B \subset A$，则 $P(A - B) = P(A) - P(B)$

推论：

1. 对任意两个随机事件 A, B，若 $B \subset A$，则 $P(A) \geq P(B)$
2. 减法公式：对任意的两个随机事件 A, B，有 $P(A - B) = P(A) - P(AB)$
3. 加法公式：对任意的两个随机事件 A, B，有 $P(A \cup B) = P(A) + P(B) - P(AB)$

条件概率

设 A, B 是两个随机事件，且 $P(B) > 0$ ，称 $P(A \mid B) = \dfrac{P(AB)}{P(B)}$ 为事件 B 发生的条件下事件 A 发生的条件概率

乘法公式

设 A, B 是两个随机事件，且 $P(B) > 0$ ，则 $P(AB) = P(B)P(A \mid B)$

同理，若 $P(A > 0)$，则 $P(AB) = P(A)P(B \mid A)$

n 元乘法公式：对于 n 个随机事件 $A_1, A_2, A_3, \cdots, A_n$ ，则

$$P(A_1A_2A_3 \cdots A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2) \cdots P(A_n \mid A_1A_2 \cdots A_{n-1})$$

全概率公式

设 $A_1,A_2,\cdots,A_n$ 是一个完备事件组，$\forall i\in[1,n]\cap Z\to P(A_i)>0$ ，则对于事件 B，有：

$$P(B)=\sum_{i=1}^{n}P(A_i)P(B\mid A_i)$$

贝叶斯公式

设 $A_1,A_2,\cdots,A_n$ 是一个完备事件组，$\forall i\in[1,n]\cap Z\to P(A_i)>0$ ，则在 B 已经发生的条件下，A_j 发生的条件概率为

$$P(A_j\mid B)=\frac{P(A_jB)}{P(B)}=\frac{P(A_j)P(B\mid A_j)}{\sum\limits_{i=1}^{n}P(A_i)P(B\mid A_i)}$$

事件的独立性

对任意两个事件 A,B ，若有 $P(AB)=P(A)P(B)$，则称事件 A 与事件 B 相互独立

若事件 A 与事件 B 独立，则 $\bar{A}$ 与 B，A 与 $\bar{B}$，$\bar{A}$ 与 $\bar{B}$ 是三组相互独立的事件

对于任意三个事件 A,B,C，若有：$\begin{cases}P(AB)=P(A)P(B)\\P(AC)=P(A)P(C)\\P(BC)=P(B)P(C)\end{cases}$

则称事件 A,B,C 两两独立，若还有 $P(ABC)=P(A)P(B)P(C)$， 则称为相互独立

试验的独立性

设 $E_1,E_2,\cdots,E_n$ 为 n 次随机试验，如果 E_1 的任一事件、E_2 的任一事件、$\cdots$、E_n 的任一事件之间都是相互独立的，则称 $E_1,E_2,\cdots,E_n$ 相互独立。

如果这 n 次独立试验是相同的，则称其为独立重复试验。

如果这 n 次试验都是只考虑一个事件 A 的发生与不发生，且 A 在每次试验中发生的概率都恒为 p ，则这类试验为伯努利概型试验。

随机变量与分布

分布函数

设 X 是随机变量，对任意实数 x，令 $F(x)=P(X\le x),x\in R$ ，则称函数 $F(x)$ 为随机变量 X 的分布函数，或记为 $F_X(x)$

性质：

1. 非负有界性：$0\le F(x)\le 1\wedge F(-\infty)=0\wedge F(+\infty)=1$
2. 单调不减性：$x_1<x_2\to F(x_1)\le F(x_2)$
3. 右连续性：$F(x+0)=F(x)$

离散型随机变量

设离散型随机变量 X 的所有可能取值为 $x_k(k=1,2,\cdots)$ ，X 取各个值的概率为 $P(X=x_k)=p_k,k=1,2,\cdots$

该式为随机变量 X 的概率分布或分布律

性质：

1. 非负性： $p_k\geq 0,k=1,2,\cdots$
2. 归一性： $\sum\limits_k p_k=1$

连续型随机变量

设随机变量 X 的分布函数为 $F(x)$ ，若存在非负实函数 $f(x)$ 使对任意的实数 x，都有 $F(x)=P(X\leq x)=\int_{-\infty}^x f(t)\mathrm{d}t$

则称 X 为连续型随机变量，其中 $f(x)$ 为 X 的概率密度函数，简称概率密度或分布密度，或记为 $f_X(x)$

性质：

1. 非负性： $f(x)\geq 0$
2. 归一性： $\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$
3. 对任意实数 $a,b(a<b)$ ，有 $P(a<X\leq b)=F(b)-F(a)=\int_a^b f(x)\mathrm{d}x$
4. $P(X=x)=F(x)-F(x-0)=0$
5. 在 $f(x)$ 的连续点 x 处，有 $F'(x)=f(x)$

正态分布的性质

$X\sim N(0,1)\rightarrow \varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

$\Phi(x)=\int_{-\infty}^x\varphi(t)\mathrm{d}t$

$X\sim N(\mu,\sigma^2)\rightarrow F(x)=\Phi(\frac{x-\mu}{\sigma})$

随机变量函数的分布

设给定离散型随机变量 X 的分布，则 $Y=g(X)$ 也是一个离散型随机变量，其所有可能的取值为 $y_k=g(x_k),k=1,2,\cdots$

$P(Y=y_k)=\sum\limits_{g(x_i)=y_k}P(X=x_i)$

已知连续型随机变量 X 的概率密度为 $f_X(x),y=g(x)$ 是连续型实函数，则求解 $f_Y(y)$ 步骤如下：

1. 求解 $F_Y(y)=P(Y\leq y)=P(g(X)\leq y)$
2. 利用 $f_Y(y)=F'_Y(y)$ 求解

多维随机变量及其分布

联合分布函数

设 (X,Y) 是二维随机变量，对于任意实数 x,y ，称二元函数 $F(x,y)=P(X\le x,Y\le y)$ 为二维随机变量 (X,Y) 的联合分布函数，简称分布函数

性质：

1. $F(x,y)$ 分别是变量 x 和变量 y 的单掉不减函数
2. $F(x,y)$ 非负有界，且对于任意固定的 x ，有 $F(x,-\infty)=0$；对任意固定的 y ，有 $F(-\infty,y)=0$ ；且 $F(-\infty,-\infty)=0,F(+\infty,+\infty)=1$
3. $F(x,y)$ 分别是变量 x,y 的右连续函数
4. 对任意的 $(x_1,y_1),(x_2,y_2)$ 若 $x_1<x_2,y_1<y_2$ ，则 $0\le F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)\le 1$

$F_X(x)=F(x,+\infty),F_Y(y)=F(+\infty,y)$ 分别为二位随机变量 (X,Y) 关于 X 和 Y 的边缘分布函数

若对于任意实数 x,y，事件 $X\le x,Y\le y$ 相互独立，即 $P(X\le x,Y\le y)=P(X\le x)P(Y\le y)$ 或 $F(x,y)=F_X(x)F_Y(y)$ 则称随机变量 X 与 Y 相互独立

二维离散型随机变量的分布

设二维随机变量 (X,Y) 的所有可能取值只有有限个或可列个，则 (X,Y) 是二维离散型随机变量。若 (X,Y) 的所有可能取值为 $(x_i,y_j),i,j=1,2,\cdots$ 且 (X,Y) 取各个可能值的概率为 $P(X=x_i,Y=y_j)=p_{ij},i,j=1,2,\cdots$

则该式称为二维离散型随机变量 (X,Y) 的联合分布律或联合分布列，简称分布律或分布列

直观上，会以表的形式给出

性质：

1. 非负性： $p_{ij}\ge 0,i,j=1,2,\cdots$
2. 归一性： $\sum\limits_i\sum\limits_j p_{ij}=1$
3. $F(x,y)=\sum\limits_{x_i\le x}\sum\limits_{y_j\le y}p_{ij}$

$P(X=x_i)=\sum\limits_j p_{ij}=p_{i\cdot},P(Y=y_j)=\sum\limits_i p_{ij}=p_{\cdot j}$ 称为二维离散型随机变量 (X,Y) 分别关于 X,Y 的边缘分布律

若 (X,Y) 为离散型随机变量，X 和 Y 相互独立的充要条件是 $P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)$ 即 $p_{ij}=p_{i\cdot}p_{\cdot j},i,j=1,2,\cdots$

若 (X,Y) 为离散型随机变量，其联合分布率为 $P(X=x_i,Y=y_j)=p_{ij},i,j=1,2,\cdots$ 。对于固定的 j ，若 $p_{\cdot j}>0$，则在条件 $Y=y_j$ 下，随机事件 $X=x_i$ 发生的概率 $P(X=x_i\mid Y=y_j)=\dfrac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\dfrac{p_{ij}}{p_{\cdot j}}$ ，称为在条件 $Y=y_j$ 下，随机变量 X 的条件分布

同理，$P(Y=y_j \mid X=x_i)=\dfrac{p_{ij}}{p_{i\cdot}}, p_{i\cdot}>0 \wedge i,j=1,2,\cdots$ ，称为在条件 $X=x_i$ 下，随机变量 Y 的条件分布

二维连续型随机变量的分布

设二维随机变量 (X,Y) 的分布函数为 $F(x,y)$，如果存在非负函数 $f(x,y)$ 使得对任意的实数 x,y 都有 $F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\mathrm{d}s\mathrm{d}t$ 则称 (X,Y) 为二维连续型随机变量。其中 $f(x,y)$ 称为 (X,Y) 的联合概率密度函数，简称概率密度或分布密度。

性质：

1. 非负性：$f(x,y)\geq 0$
2. 归一性：$\int_{-\infty}^{+\infty}\int_{\infty}^{+\infty} f(s,t)\mathrm{d}s\mathrm{d}t=F(+\infty,+\infty)=1$
3. 若 $f(x,y)$ 在点 (x,y) 连续，则 $\dfrac{\partial^2 F(x,y)}{\partial x\partial y}=\dfrac{\partial^2 F(x,y)}{\partial y\partial x}=f(x,y)$
4. 对任意 $(x_1,y_1),(x_2,y_2)$ 且 $x_1<x_2,y_1<y_2$，则 (X,Y) 落在矩形 $(x_1,x_2]\times(y_1,y_2]$ 内的概率为 $P(x_1<X\leq x_2,y_1\leq Y\leq y_2)=\int_{x_1}^{x_2}\int_{y_1}^{y_2} f(s,t)\mathrm{d}t\mathrm{d}s$

设 G 为平面上的有界区域，若二维随机变量 (X,Y) 的分布密度函数为 $f(x,y)=\begin{cases}\frac{1}{S_G},(x,y)\in G\\ 0,\text{其他}\end{cases}$

其中 $S_G=\iint_G \mathrm{d}x\mathrm{d}y$ 为区域 G 的面积，则称二维随机变量 (X,Y) 服从 G 上的均匀分布

二维正态分布的记号 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$

其中 μ_i 和 σ_i^2 分别表示第 i 个变量的期望、方差；ρ 表示相关系数

若 (X,Y) 为连续型随机变量，其联合概率密度为 $f(x,y)$，边缘密度为 $f_X(x),f_Y(y)$，则 X 和 Y 相互独立的充要条件是 $f(x,y)=f_X(x)f_Y(y)$ 在 $f(x,y),f_X(x),f_Y(y)$ 的一切公共连续点上都成立

设 (X,Y) 为二维连续型随机变量，联合密度和关于 X,Y 的边缘分布密度分别为 $f(x,y),f_X(f),f_Y(y)$，对于固定的 x，若 $f_X(x)>0$，则称 $f_{Y|X}(y\mid x)=\frac{f(x,y)}{f_X(x)}$ 为在条件 $X=x$ 下，随机变量 Y 的条件概率密度

类似定义 $f_{X|Y}(x\mid y)$

二维随机变量极值函数的分布

设随机变量 X 与 Y 相互独立，其分布函数分别为 $F_X(x),F_Y(y)$，则极值 $M=\max X,Y,N=\min X,Y$ 的分布函数分别为：

$F_M(x)=P(X\leq x,Y\leq x)=F_X(x)F_Y(x)$

$F_N(x)=1-P(X>x,Y>x)=1-[1-F_X(x)][1-F_Y(y)]$

正态分布的可加性

对于随机变量 $X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)$ ，有 $aX+bY\sim N(a\mu_1+b\mu_2,a^2\sigma_1^2+b^2\sigma_2^2)$

对于随机变量 $X_i \sim N(\mu_i, \sigma_i^2)$ 和常数 $a_i \in R$

有 $\sum\limits_i a_i X_i \sim N(\sum\limits_i a_i \mu_i, \sum\limits_i a_i^2 \sigma_i^2)$

随机变量的数学特征

数学期望

设离散型随机变量 X 的分布律为 $P(X = x_k) = p_k, k = 1, 2, \cdots$ ，若级数 $\sum\limits_k x_k p_k$ 绝对收敛，则称级数 $\sum\limits_k x_k p_k$ 的和为离散型随机变量 X 的数学期望，记为 $E(X)$ 或 EX，即 $E(X) = \sum\limits_k x_k p_k$

设连续型随机变量 X 的概率密度为 $f(x)$，若 $\int_{-\infty}^{+\infty} x f(x) \mathrm{d}x$ 绝对收敛，则称级数 $\int_{-\infty}^{+\infty} x f(x) \mathrm{d}x$ 的和为离散型随机变量 X 的数学期望，记为 $E(X)$ 或 EX，即 $E(X) = \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x$

随机变量函数的数学期望：将 x 改为 $g(x)$ ：

离散型： $E(g(x)) = \sum\limits_k g(x_k) p_k$

连续型： $E(g(x)) = \int_{-\infty}^{+\infty} g(x) f(x) \mathrm{d}x$

性质：

1. 设 X 为随机变量，则对任意常数 a, b 有 $E(aX + b) = aE(X) + b$
2. 设 b 为常数，则 $E(b) = b$
3. 数学期望的线性性质：设 X, Y 为两个随机变量，a, b 为任意常数，有 $E(aX + bY) = aE(X) + bE(Y)$

> 设 $X_1, X_2, \cdots, X_n$ 为 n 个随机变量，$a_1, a_2, \cdots, a_n$ 为 n 个常数，有 $E(\sum\limits_{i=1}^n a_i X_i) = \sum\limits_{i=1}^n a_i E(X_i)$

4. 设随机变量 X, Y 相互独立且数学期望均存在，则 $E(XY) = E(X)E(Y)$

方差

设 X 为随机变量，若 $E[X - E(X)]^2$ 存在，则称 $E[X - E(X)]^2$ 为 X 的方差，记为 $D(X)$ 或 DX，即 $D(X) = E[X - E(X)]^2$ 。

同时称 $\sqrt{D(X)}$ 为 X 的标准差或均方差，记为 $\sigma(X)$ ，即 $\sigma(X) = \sqrt{D(X)}$ 。

方差的计算式：$D(X) = E(X^2) - [E(X)]^2$

性质：

1. 设 X 为随机变量，则对于任意常数 a, b，有 $D(aX + b) = a^2 D(X)$

> $D(X) = D(-X)$

2. 设 b 为常数，则 $D(b) = 0$
3. 设 $X_1, X_2, \cdots, X_n$ 为期望和方差均存在的随机变量，则
$$\begin{aligned} D(\sum_{i=1}^{n} X_i) &= \sum_{i=1}^{n}\sum_{j=1}^{n} E(X_iX_j) - E(X_i)E(X_j) \\ &= \sum_{i=1}^{n} D(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E(X_iX_j) - E(X_i)E(X_j) \end{aligned}$$

> $n = 2$ 时，有
> $D(X+Y) = D(X) + D(Y) + 2E(XY) - E(X)E(Y) = D(X) + D(Y) + 2 \cdot \mathrm{cov}(X, Y)$

4. 设 X, Y 为相互独立的随机变量，则 $D(X+Y) = D(X) + D(Y)$
5. 设 $X_1, X_2, \cdots, X_n$ 为期望和方差均存在的相互独立的随机变量，且 $a_1, a_2, \cdots, a_n$ 为常数，则
$$D(\sum_{i=1}^{n} a_iX_i) = \sum_{i=1}^{n} a_i^2 D(X_i)$$

对任何随机变量 X， 若它的数学期望 $E(X)$ 和方差 $D(x)$ 都存在，且 $D(X) > 0$，则称 $X^* = \frac{X-E(X)}{\sqrt{D(X)}}$ 为 X 的标准化随机变量

协方差与相关系数

设 (X, Y) 为二维随机变量，若 $E[X - E(X)][Y - E(Y)]$ 存在，则称它是随机变量 X 与 Y 的协方差，记为 $\mathrm{cov}(X, Y)$

特殊的，有 $\mathrm{cov}(X, X) = D(X)$

计算式：$\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y)$

性质：

1. $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$
2. 若 a 为常数，则 $\mathrm{cov}(X, a) = 0$
3. 若 a, b 为常数，则 $\mathrm{cov}(aX, bY) = ab \cdot \mathrm{cov}(X, Y)$
4. $\mathrm{cov}(X_1 + X_2, Y) = \mathrm{cov}(X_1, Y) + \mathrm{cov}(X_2, Y)$
5. 若对二维随机变量 X, Y ，有 $D(X \pm Y) = D(X) + D(Y) \pm 2 \cdot \mathrm{cov}(X, Y)$

若 $\mathrm{cov}(X, Y) = 0$ 则称 X, Y 不相关； 若 $\mathrm{cov}(X, Y) < 0$ 则称 X, Y 负相关； 若 $\mathrm{cov}(X, Y) > 0$ 则称 X, Y 正相关。

对任意 n 个随机变量 $X_1, X_2, \cdots, X_n$ ，有：

$$D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} D(X_i) + 2\sum_{1 \le i < j \le n} \mathrm{cov}(X_i, X_j)$$

设 (X, Y) 是二维随机变量，若 $D(X) > 0, D(Y) > 0$，则称 $\frac{\mathrm{cov}(X,Y)}{\sqrt{D(X)D(Y)}}$ 为随机变量 X 与 Y 的相关系数，记为 ρ 或 ρ_{XY}，即 $\rho_{XY} = \dfrac{\mathrm{cov}(X, Y)}{\sqrt{D(X)D(Y)}}$

同时，还有 $\rho_{XY}=E(\frac{X-E(X)}{\sqrt{D(X)}}\cdot\frac{Y-E(Y)}{\sqrt{D(Y)}})=E(X^*\cdot Y^*)$

性质：

1. $|\rho_{XY}|\le 1$
2. $|\rho_{XY}|=1$ 的充要条件是存在常数 $a\neq 0,b$ 使得 $P(Y=aX+b)=1$，即随机变量 X 与 Y 以概率 1 有线性关系

二维正态分布中，$\rho=0$ 时，X 与 Y 相互独立

X 与 Y 相互独立 $\Rightarrow$ X 与 Y 不相关 $\Leftrightarrow \rho_{XY}=0 \Leftrightarrow \mathrm{cov}(X,Y)=0 \Leftrightarrow E(XY)=E(X)E(Y)$ $\Leftrightarrow D(X\pm Y)=D(X)+D(Y)$

极限定理初步

依概率收敛

设 X_n 为随机变量序列，X 为随机变量，如果对任意的 $\varepsilon>0$ 有 $\lim\limits_{n\to\infty}P(|X_n-X|<\varepsilon)=1$ ，则称 X_n 依概率收敛于 X ，记做 $X_n\xrightarrow{P}X$

大数定律的一般形式

设 X_n 为随机变量序列，如果 $E(X_n)(n=1,2,\cdots)$ 存在，使得对任意 $\varepsilon>0$ 有 $\lim\limits_{n\to\infty}P|\frac{1}{n}\sum\limits_{i=1}^{n}X_i-\frac{1}{n}\sum\limits_{i=1}^{n}E(X_i)|<\varepsilon=1$ ，即 $\frac{1}{n}\sum\limits_{i=1}^{n}X_i-\frac{1}{n}\sum\limits_{i=1}^{n}E(X_i)\xrightarrow{P}0$ ，则称随机变量序列 X_n 服从大数定律

> 样本均值依概率收敛于期望，即 $\frac{1}{n}\sum\limits_{i=1}^{n}X_i\xrightarrow{P}E(\frac{1}{n}\sum\limits_{i=1}^{n}X_i)$

伯努利大数定律

1. 设 X 是 n 次伯努利试验中随机事件 A 发生的次数，p 是每次试验时事件 A 发生的概率
2. 则对任何 $\varepsilon>0$，有 $\lim\limits_{n\to\infty}P(|\frac{X}{n}-p|<\varepsilon)=1$ ，或 $\frac{X}{n}\xrightarrow{P}p$

切比雪夫大数定律

1. 设 X_n 为两两不相关的随机变量序列
2. 若每个随机变量 X_i 的方差存在，且有共同的上街，即 $D(X_i)\le c(c\text{ 为常数 }),i=1,2,\cdots$
3. 则随机变量序列 X_n 服从大数定律，即对任何 $\varepsilon>0$ 有
   $\lim\limits_{n\to\infty}P(|\frac{1}{n}\sum\limits_{i=1}^{n}X_i-\frac{1}{n}\sum\limits_{i=1}^{n}E(X_i)|<\varepsilon)=1$ 或 $\frac{1}{n}\sum\limits_{i=1}^{n}X_i-\frac{1}{n}\sum\limits_{i=1}^{n}E(X_i)\xrightarrow{P}0$

辛钦大数定律

1. 设 X_n 为独立同分布随机变量序列

2. 若每个随机变量 X_i 的数学期望存在，即 $E(X_i)=\mu(i=1,2,\cdots)$
3. 则随机变量序列 X_n 服从大数定律，即对任何 $\varepsilon>0$ 有 $\lim\limits_{n\to\infty}P|\frac{1}{n}\sum\limits_{i=1}^{n}X_i-\mu|<\varepsilon=1$ 或 $\frac{1}{n}\sum\limits_{i=1}^{n}X_i\xrightarrow{P}\mu$

独立同分布中心极限定理

1. 设随机变量 $X_1,X_2,\cdots,X_n,\cdots$ 相互独立同分布
2. 且数学期望和方差存在：$E(X_i)=\mu,D(X_i)=\sigma^2\neq0,(i=1,2,\cdots)$
3. 则随机变量 $\sum\limits_{i=1}^{n}X_i\dot\sim N(n\mu,n\sigma^2)(n\to+\infty)$

二项分布中心极限定理

1. 设 X 为 n 次伯努利试验中事件 A 出现的次数，$p(0<p<1)$ 是每次试验中事件 A 发生的概率，即 $X\sim B(n,p)$
2. 则随机变量 $\sum\limits_{i=1}^{n}X_i\dot\sim N(np,np(1-p))$

数理统计的基本概念与抽样分布

总体与样本

1. 统计学中，把研究问题所设计的对象的全体称为总体
2. 把总体中的每个成员称为个体
3. 每次抽取 n 个个体，这 n 个个体 $X_1,X_2,\cdots,X_n$ 就称为总体 X 的一个容量为 n 的样本或子样
4. 其中样本所包含的个体数量 n 称为样本容量或样本大小

设 $X_1,X_2,\cdots,X_n$ 为总体 X 的一个容量为 n 的样本，若满足：

1. 独立性：即 $X_1,X_2,\cdots,X_n$ 相互独立
2. 同分布性：即每一个 $x_i(i=1,2,\cdots,n)$ 都与总体 X 服从相同的分布，则这样的样本称为简单随机样本，简称样本

统计量

设 $X_1,X_2,\cdots,X_n$ 是总体 X 的样本，$T(X_1,X_2,\cdots,X_n)$ 是样本的实值函数，且不包含任何位置参数，则称 $T(X_1,X_2,\cdots,X_n)$ 为统计量

1. 统计量 $\bar{X}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$ 称为样本的均值
2. 统计量 $S^2=\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n-1}[\sum\limits_{i=1}^{n}X_i^2-n\bar{X}^2]$ 称为样本的方差
3. 统计量 $S=\sqrt{S^2}$ 称为样本的均方差或标准差

4. 统计量 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$ 称为样本的 k 阶原点矩$(k=1,2,\cdots)$
5. 统计量 $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$ 称为样本的 k 阶中心矩$(k=1,2,\cdots)$

设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，对应的顺序统计量为 $X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}$

6. 统计量 $R=X_{(n)}-X_{(1)}$ 称为样本的极差
7. 统计量 $M_e=\begin{cases}X_{(\frac{n+1}{2})},n\text{ 是奇数}\\ \frac{1}{2}(X_{(\frac{n}{2})}+X_{(\frac{n}{2}+1)}),n\text{ 是偶数}\end{cases}$

经验分布函数

设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，对应的顺序统计量为 $X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}$ ，当给定顺序统计量的观测值 $x_{(1)}\le x_{(2)}\le\cdots\le x_{(n)}$ 时，对任意实数 x ，称下列函数：

$$F_n(x)=\begin{cases}0,x<x_{(1)}\\ \frac{k}{n},x_{(k)}\le x<x_{(k+1)}\wedge k=1,2,\cdots,n-1\\ 1,x_{(n)}\le x\end{cases}$$

为总体 X 的经验分布函数

χ^2 分布

设 $X_1,X_2,\cdots,X_n$ 为独立同分布的随机变量，且都服从 $N(0,1)$，则称随机变量 $X=X_1^2+X_2^2+\cdots+X_n^2=\sum_{i=1}^{n}X_i^2$ 所服从的分布为自由度为 n 的 χ^2 分布，记作 $X\sim\chi^2(n)$

性质：

1. 可加性：设 $Y_1\sim\chi^2(m),Y_2\sim\chi^2(n)$，且 Y_1,Y_2 相互独立，则 $Y_1+Y_2\sim\chi^2(m+n)$
2. 若 $X\sim\chi^2(n)$，则 $E(X)=n,D(X)=2n$

t 分布

设随机变量 $X\sim N(0,1),Y\sim\chi^2(n)$ 且 X,Y 相互独立，则称随机变量 $T=\dfrac{X}{\sqrt{Y/n}}$ 所服从的分布为自由度为 n 的 t 分布，记作 $T\sim t(n)$

性质：

1. 设 $T\sim t(n)(n>1)$ 则 $E(T)=0$
2. 设 $T\sim t(n)$，当 n 较大时，$T\dot{\sim}N(0,1)$

F 分布

设随机变量 $X\sim\chi^2(m),T\sim\chi^2(n)$ ，且 X,Y 相互独立，则称随机变量 $F=\frac{X/m}{Y/n}$ 所服从的分布为自由度为 m,n 的 F 分布，记作 $F\sim F(m,n)$

性质：

1. 若 $X \sim F(m,n)$ ，则 $\frac{1}{X} \sim F(n,m)$
2. 若 $X \sim t(n)$，则 $X^2 \sim F(1,n)$

概率分布的分位点

x_α 为概率的上 α 分位点，当且仅当 $P(X > x_\alpha) = \alpha$

1. 正态分布的分位点满足 $u_{1-\alpha} = -u_\alpha$
2. $t_{1-\alpha}(n) = -t_\alpha(n)$
3. $F_\alpha(m,n) = \dfrac{1}{F_{1-\alpha}(n,m)}$

正态总体的抽样分布

设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $X \sim N(\mu, \sigma^2)$ 的样本，则

1. $\bar{X} \sim N(\mu, \dfrac{\sigma^2}{n})$
2. $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
3. $\bar{X}$ 与 S^2 相互独立
4. $\dfrac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1)$

推论：

1. $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
2. $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$
3. $\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2 = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
4. $\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$

设 $X_1, X_2, \cdots, X_m$ 是来自正态总体 $X \sim N(\mu_1, \sigma_1^2)$ 的样本，$Y_1, Y_2, \cdots, Y_n$ 是来自正态总体 $Y \sim N(\mu_2, \sigma_2^2)$ 的样本，且 X, Y 相互独立

1. $\bar{X} - \bar{Y} \sim N(\mu_1-\mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n})$
2. $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(m-1, n-1)$

参数估计

矩法估计

1. 计算总体分布的 r 阶原点矩 $E(X^r)$，记为 $E(X^r)=g_r(\theta),r=1,2,\cdots,k$
2. 近似替换，即用样本 r 阶原点矩替换成总体 r 阶原点矩，列出方程 $g_r(\theta)\triangleq\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^r(r=1,2,\cdots,k)$
3. 解此方程得 $\theta=h(X_1,X_2,\cdots,X_n)$

则以 $h(X_1,X_2,\cdots,X_n)$ 作为 θ 的估计量 $\hat{\theta}$，并称 $\hat{\theta}=h(X_1,X_2,\cdots,X_n)$ 为 θ 的矩估计量，而称 $h(X_1,X_2,\cdots,X_n)$ 为 θ 的矩法估计值

极大似然估计

设总体 X 的概率分布为 $f(x;\theta)$，其中 $\theta=(\theta_1,\theta_2,\cdots,\theta_k)$ 是未知量数 $X_1,X_2,\cdots,X_n$ 是总体 X 的样本，则称 $X_1,X_2,\cdots,X_n$ 的联合概率分布

$$L(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n}f(x_i;\theta)$$

为样本的似然函数，简记为 $L(\theta)$

若有 $\hat{\theta}\in\Theta$，使得 $L(\hat{\theta})=\sup\limits_{\theta\in\Theta}L(\theta)$，则称 $\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$ 为 θ 的极大似然估计值，而称对应的统计量 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为 θ 的极大似然估计量

一般步骤如下：

1. 根据总体 X 的分布 $f(x;\theta)$ 写出似然函数 $L(\theta)$
2. 直接求解 $L(\theta)$ 最大值；或对似然函数取对数 $\ln L(\theta)=\sum\limits_{i=1}^{n}\ln f(x_i;\theta)$ ，求解似然方程 $\dfrac{\mathrm{d}\ln L}{\mathrm{d}\theta}=0$

估计量的评价标准

设 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是未知参数 θ 的估计量，$\theta\in\Theta$。若 $E(\hat{\theta})=\theta$，则称 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是 θ 的无偏估计。如果 $E(\hat{\theta})\neq\theta$，则 $E(\hat{\theta})-\theta$ 称为估计量 $\hat{\theta}$ 的偏差

> $E(S)=\sigma^2$

若 $\lim\limits_{n\to\infty}E(\hat{\theta})=\theta$，则称 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是 θ 的渐进无偏估计量

设 $\hat{\theta}_1(X_1,X_2,\cdots,X_n),\hat{\theta}_2(X_1,X_2,\cdots,X_n)$ 均是参数 θ 的两个无偏估计量，若 $D(\hat{\theta}_1)<D(\hat{\theta}_2)$ 则称估计量 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效

设 $\hat{\theta}_n$ 是 θ 的估计量，若对任意的 $\varepsilon>0$，有 $\lim\limits_{n\to\infty}P(|\hat{\theta}_n-\theta|<\varepsilon)=1$ 恒成立，则称 $\hat{\theta}_n$ 是 θ 的一致估计量，也可以说估计量 $\hat{\theta}_n$ 具有一致性，记为 $\hat{\theta}_n\xrightarrow{P}\theta$

区间估计

设总体 X 的分布函数是 $F(x;\theta)$，其中 θ 是未知参数，$X_1,X_2,\cdots,X_n$ 是从总体 X 中抽取的样本，对于给定的 $\alpha(0<\alpha<1)$，构造统计量 $\hat{\theta}_1(X_1,X_2,\cdots,X_n),\hat{\theta}_2(X_1,X_2,\cdots,X_n)$ 。如果满足 $P(\hat{\theta}_1\le\theta\le\hat{\theta}_2)=1-\alpha$ ，则称随机区间 $(\hat{\theta}_1,\hat{\theta}_2)$ 为参数 θ 的置信度为 $1-\alpha$ 的双侧置信区间（简称置信区间），其中 $\hat{\theta}_1,\hat{\theta}_2$ 分别称为置信下限和置信上限，$1-\alpha$ 称为置信度或置信水平。

单个正态总体参数的置信区间

1. 总体方差 σ^2 已知时，总体均值 μ 的置信区间：

由 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

μ 的双侧置信区间为 $(\bar{X}-u_{\alpha/2}\frac{\sigma}{\sqrt{n}},\bar{X}+u_{\alpha/2}\frac{\sigma}{\sqrt{n}})$

2. 总体方差 σ^2 未知时，总体均值 μ 的置信区间：

由 $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

μ 的双侧置信区间为 $(\bar{X}-t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}},\bar{X}+t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}})$

3. 总体方差 μ 已知时，总体均值 σ^2 的置信区间：

由 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$

σ^2 的双侧置信区间为 $(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)},\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)})$

σ 的双侧置信区间为 $(\sqrt{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)}},\sqrt{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)}})$

4. 总体方差 μ 未知时，总体均值 σ^2 的置信区间：

由 $\frac{n-1}{\sigma^2}S^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 \sim \chi^2(n-1)$

σ^2 的双侧置信区间为 $(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)})$

σ 的双侧置信区间为 $(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}},\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}})$

置信区间长度与置信度的关系

样本容量一定时，置信区间长度 l 的变小与置信度 $(1-\alpha)$ 的增大不可能同时成立。

证明：（以 σ^2 已知时，μ 的双侧置信区间为例）

$(1-\alpha)$ 增大时，α 减小，$u_{\alpha/2}$ 增大

双侧置信区间长度 $l=2\cdot\frac{\sigma}{\sqrt{n}}u_{\alpha/2}$ 增大

同理，仅当 $(1-\alpha)$ 减小时，l 变小

假设检验

假设检验的基本思想

关于总体的假设通常的两个相互对立的假设，我们把需要检验是否为真的假设称为原假设或零假设，用 H_0 表示，与之对立的另一个假设称为备择假设或对立假设，用 H_1 表示。

假设检验的基本思想是带有某种概率性质的反证法，基于"小概率事件"原理。

> "小概率事件原理"：一个事件如果发生的概率很小的话，那么它在一次试验中是几乎不可能发生的

构造小概率事件 A：选一个较小的概率 α 称之为检验的显著水平，寻找 a,b 使得检验统计量 T 满足 $P((T_0<a)\cup T_0>b))\le\alpha$

在 H_0 成立的条件下，事件 $A=(T_0<a)\cup(T_0>b)$ 是小概率事件

通常把拒绝原假设的区域 W 称为关于原假设 H_0 的拒绝域，简称拒绝域；而把接受原假设的区域称为关于原假设 H_0 的接受与，简称接受域。把区间端点称为临界值。

第一类错误："弃真"错误：由于样本的随机性，检验统计量的观测值落入了拒绝域，从而拒绝了正确的原假设 H_0。其概率 $P($ 拒绝 $H_0\mid H_0$ 为真 $)=\beta_1$

第二类错误："取伪"错误：由于样本的随机性，检验统计量的观测值落入了接受域，从而接收了错误的原假设 H_0。其概率 $P($ 接受 $H_0\mid H_0$ 为假 $)=\beta_2$

当样本容量 n 固定时，要使 β_1,β_2 同时变小是不可能的

当样本容量 n 固定时，着重控制犯第一类错误的概率，使之不超过某一给定值 α，称之为显著性水平，即 $P($ 拒绝 $H_0\mid H_0$ 为真 $)=\beta_1\le\alpha$

单个正态总体参数的假设检验

设给定的显著水平为 α

1. 总体方差 σ^2 已知时，总体均值 μ 的假设检验

由 $T_0=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

拒绝域为 $W=|T_0|>u_{\alpha/2}$

> U 检验

2. 总体方差 σ^2 未知时，总体均值 μ 的假设检验

由 $T_0=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

拒绝域为 $W=|T_0|>t_{\alpha/2}(n-1)$

> t 检验

3. 总体均值 μ 已知时，总体方差 σ^2 的假设检验

由 $T_0 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2(n)$

拒绝域为 $W = |T_0| > \chi^2_{\alpha/2}(n-1)$

> χ^2 检验

4. 总体均值 μ 未知时，总体方差 σ^2 的假设检验

由 $T_0 = \frac{n-1}{\sigma^2}S^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1)$

拒绝域为 $W = |T_0| > \chi^2_{\alpha/2}(n)$

> χ^2 检验