基本概念/一些公式

2019年11月20日　　19:14

$1 - P(B) = P(\bar{B})$

$P(AUB) = P(A) + P(B) - P(AB)$

注：括号里的和并集用∪表示，不用+。

用$\binom{n}{r}$表示C_n^r。（可以记：上面的n是总数，掉下来的一部分r就是被选择的数量）

独立性： $\mathrm{P(AB)} = \mathrm{P}(A)\mathrm{P(B)}$　不是随时都成立的

条件概率：　$P(B|A) = \frac{P(AB)}{P(A)}$。

　　左边表示在A事件发生的基础上，B事件发生的概率。

　　右边分子表示AB同时发生，分母表示A发生概率。

　　记忆：相当于左边括号里的从躺着变成站起来，再乘一下。

　　基本概率公式，依然通用。

全概率公式：　$P(A) = \sum_{i \in I} \cancel{P(A|B_i)} P(B_i)$　$\frac{A}{B_i} \times B_i$

　　　记忆：右侧第一项想成分数形式，那就显而易见了。

　　　应用：由原因推结果

将S划分成若干个集合，且互相无交集，则称事件组是一个**完备事件组。**

贝叶斯公式：　$$\mathrm{P}(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j \in I} P(A|B_j)P(B_j)}$$

　　应用：由结果推原因

伯努利概型： $P(B_k) = \binom{n}{k} p^k q^{n-k}$　$(0 \le k \le 4)$

　　P指事件发生概率，q指事件不发生概率。

第二章 随机变量及其分布

2019年11月20日　　19:09

随机变量

2019年11月20日 19:09

离散型随机变量

2019年11月20日　　19:09

分布列（分布律、概率分布）

即列出所有可能结果，填表，附上每种结果发生的概率。

几种常见的离散型分布

1. **单点分布**

   只有发生概率为1的一个事件。

   称 X服从 **单点分布**或**退化分布**。

2. **两点分布**

   X只取0或1，且发生概率为p

   X 服从 **参数为p的两点分布或0-1分布**。

3. **二项分布**

   参考伯努利概型

   $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}\quad (k=0,1,2\ldots n)$

4. **泊松分布**

   当二项分布随着n无限增大时，按照上述公式计算会很麻烦。

   则，有$n\to+\infty$时有$np_n\to\lambda$，那么

   $\lim_{n\to+\infty}\binom{n}{k}p^k(1-p)^{n-k}\approx\frac{\lambda^k}{k!}e^{-\lambda}$　　（**泊松分布**）

   称 X 服从**参数为λ 泊松分布**，记作$\boldsymbol{X\sim P(\lambda)}$. $(\lambda=\boldsymbol{np})$

5. **几何分布**

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

   称X服从**参数为p的几何分布**，记为X $\sim g(p)$

   背景：事件A发生概率为p，重复k次，直到A事件发生时停止。

   无论重复多少次，下一次的概率不会变，称为**几何分布的无记忆性**。

6. **超几何分布**

   X的分布律为：

   $$P(X=m)=\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}\ (m=0,1,2,\ldots,n)$$

   称X服从**超几何分布**。

   背景：假设有N个产品，其中M个是正品，$N-M$个是次品，从中一次性地去除n个产品，则其中含有的正品数X服从**超几何分布**。

7. **幂律分布**

   若 $P(X=k)=\frac{C}{k^\gamma},(k=1,2,\ldots)$　*($\gamma>1$, C为归一化常数)*

   称X服从**参数为γ的幂律分布**。

随机变量的分布函数

2019年11月20日 19:09

分布函数

设X是一个随机变量，x是任意实数，称

$$F(X) = P(X \le x)$$

为随机X的**分布函数**，称X服从$F(x)$，记为$X \sim F(x)$.

对于任意$a < b$有

$$P(a < X < b) = F(b) - F(a)$$

性质：

1. $F(X)$单调不减
2. $0 \le F(X) \le 1; F(-\infty) = 0; F(+\infty) = 1;$
3. $F(X)$是X的右连续函数

连续性随机变量及其密度函数

2019年11月20日　　19:10

密度函数：通俗的说，就是分布函数的导数。

$$\mathrm{F(X)}=\int_{-\infty}^{x} f(t) d t$$

f(x)是随机变量X的**概率密度函数**，简称**概率密度**或**密度。**

性质：

1. 非负性：对任意x，均有$\mathrm{f(x)} \geq 0$
2. 规范性：$\int_{-\infty}^{+\infty} f(t) d t=1$

定理2.4.1：

1. $P(a \leq X \leq b)=\int_{a}^{b} f(x) \mathrm{d} x$
2. $f\left(x_{0}\right)=F'\left(x_{0}\right)$
3. $P(x=c)=0$

几种常见的分布

1. 均匀分布

   记作$\mathbf{X} \sim \boldsymbol{U}(\boldsymbol{a}, \boldsymbol{b})$。X在此区间上，均匀分布。

   $$\mathrm{f(x)}=\begin{cases}\dfrac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{其他}\end{cases}$$

2. 指数分布（唯一的无记忆型的连续型分布）

   记作$\mathbf{X} \sim \boldsymbol{E}(\boldsymbol{\lambda})$

   $$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x<0\end{cases}$$

3. 柯西分布

   $$f(x)=\frac{1}{\pi\left(1+x^{2}\right)}$$

   $$F(x)=\frac{1}{\pi} \arctan x+\frac{1}{2}$$

4. 正态分布

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

   称X服从参数为μ, σ^{2}的正态分布，记作$\mathrm{X} \sim N\left(\mu, \sigma^{2}\right)$。

正态分布计算公式及3σ原则

当$\mu=0$，$\sigma=1$时，称其位标准正态分布。

即当$\mathrm{U} \sim N(0,1)$时：

$\phi(-u) = 1 - \phi(u)$
$\mathrm{P(U > u) = 1} - \phi(u)$
$\mathrm{P(a < U < b)} = \phi(b) - \phi(a)$
$\mathrm{P(|U| < c)} = 2\phi(c) - 1$

对于一般正态分布，有

$$\mathrm{F(x)} = \phi(\frac{x - \mu}{\sigma})$$

随机变量函数分布

2019年11月20日 19:10

离散型随机变量的情形

1. 确定X的每个取值
2. 确定每一个$P(X)$

连续型随机变量的情形

已知X的分布函数$F_X(x)$和概率密度$f_X(x)$，求$Y = h(X)$的概率密度$f_Y(y)$。
思路：先求Y的分布函数，再求导得到概率密度函数。

1. 确定Y的取值范围，不属于该范围时的概率为0；
2. 当Y在取值范围内时，先求Y的分布函数。关键是把$\{Y \leq y\}$中的Y用X表示。
   就可以把未知数Y的概率用已知分布函数X来表示。

第三章 多维随机变量及其分布

2019年11月20日 19:10

二维随机变量及其分布函数

2019年11月20日　　19:10

类似于一维变量的分布函数，二维分布函数为

$F(x,y) = P(\{X \le x\} \cap \{Y \le y\}) = P(X \le x, Y \le y)$

也叫做二维随机变量(X,Y)的**分布函数**或者X与Y的**联合分布函数**。

几何意义：可以看作点(X,Y)落在点(x,y)左下方无穷矩形的概率。

边缘分布

$F_X(x) = F(x,+\infty)$ 是X的边缘分布函数。　(即$X \to x, Y \to +\infty$)

同理，有$F_Y(y) = F(y,+\infty)$ 是Y的边缘分布函数。　(即$X \to +\infty, Y \to y$)

联合分布函数可以确定唯一的边缘分布函数，反之却不成立。

联合密度函数可以确定唯一的边缘密度函数，反之却不成立。

定理3.4.2 对于服从二维正态分布$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$的二维随机变量，X和Y相互独立的充要条件是$\rho = 0$。

二维离散型随机变量及其分布律

2019年11月20日　19:10

X/Y	y_1	y_2	…	y_j	…
x_1	p_{11}	p_{12}	…	y_{1j}	…
x_2	p_{21}	p_{22}	…	p_{2j}	…
…	…	…	…	…	…
x_i	p_{i1}	p_{i2}	…	p_{ij}	…
…	…	…	…	…	…

此为xy二维变量的**分布律**或**联合分布律**。

边缘分布律

$\mathrm{P}(X = x_i) = p_{i\cdot}$　x的**边缘分布律**（第i行的概率之和）

$\mathrm{P}(Y = y_i) = p_{\cdot j}$　y的**边缘分布律**（第j列的概率之和）

照图可以很好的理解“边缘”二字。

二维连续性随机变量及其密度函数

2019年11月20日　　19:10

类似于一维变量

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv$$

称(X,Y)是**二维连续型随机变量，**称$f(x,y)$为**概率密度函数**或**联合概率密度。**

1. 非负性
2. 规范性

参考一维变量

边缘密度

对于概率密度函数积一半分。求哪个变量，就对另一个积分。

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy$$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx$$

二维变量有　均匀分布和正态分布。

随机变量的独立性

2019年11月20日　　19:11

对于变量随机X，Y

若 $F(x,y) = F_X(x)\cdot F_Y(y)$成立，则称x，y相互独立。

其他判定:

$P(X = x_i, Y = y_i) = P(X = x_i)P(Y = y_i)$

$f(x,y) = f_X(x)f_Y(y)$

若二维随机变量X，Y服从二维正态分布$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,则XY互相独立的充要条件是$\rho = 0$。

定理3.4.3 若X，Y是相互独立的随机变量，且$h(x), g(y)$均为连续或单调函数，则$h(X)$和$g(Y)$也是相互独立。

条件分布

2019年11月20日 19:11

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

是 在给定条件Y = y 下，随机变量X的**条件密度函数。**

有概率密度函数后，便可求分布函数。

还是要掌握求二维变量下的边缘密度分布函数（上式中的分母）。

对于离散型，列表求。

对于连续性，公式求。（积分）详细的看 边缘密度函数

n维随机变量

2019年11月20日 19:11

由二维扩展到n维
大多类似扩展即可。

联合分布；
独立性；

两个随机变量的函数分布

2019年11月20日　19:11

1. 二项分布
2. 泊松分布
3. 正态分布　有限个独立正态随机变量的线性组合仍为正态随机变量

均有可加性。

$X \sim B(n_1,p)$　$Y \sim B(n_2,p)$,则 $X+Y \sim B(n_1+n_2,p)$

$X \sim P(\lambda_1)$ $Y \sim P(\lambda_2)$,则 $X+Y \sim P(\lambda_1+\lambda_2)$

$X \sim N(\mu_1,\sigma_1^2)$ $Y \sim N(\mu_2,\sigma_2^2)$,则$X+Y \sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$

标记和边缘密度函数一样，但这里就是单个变量的密度函数。这样记是为了区分x和y。

设X Y是两个相互独立的随机变量，并且知道它们的密度函数$f_X(x)$ $f_Y(y)$，求$Z=X+Y$的密度函数。

想法类似于一维变量的求法 随机变量函数的分布

直接求密度函数很难，同样地，先求出它的分布函数，再通过求导求密度函数。

先弄清几个式子。　因为xy是相互独立的。所以$f(x,y) = f(x)f(y)$

$$F_Z(z) = P(Z \le z) = P(X+Y \le z) = \iint_{x+y\le z} f(x,y)dxdy = \iint_{x+y\le z} f(x)f(y)dxdy$$

然后就是二重积分的计算了。上式又叫做卷积公式。

定理3.7.4 最大值和最小值的分布

设$X_1, X_2, \dots, X_n$ 相互独立，M，N分别为所有变量的最大值和最小值函数。

M的分布函数为

$$F_M(x) = F_{X_1}(x) \dots F_{X_2}(x)$$

N的分布函数为

$$F_N(x) = 1 - [1 - F_{X_1}(x)] \dots [1 - F_{X_2}(x)]$$

计算离散型随机变量和连续型随机变量的函数的分布：

利用全概率公式，和条件概率

在每个离散型条件下发生概率的和。

第四章 随机变量的数字特征

2019年11月20日 19:11

随机变量的数学期望

2019年11月20日 19:11

$$E(x) = \sum_{k=1}^{+\infty} x_k p_k$$

数学期望就是随机变量的取值以它们的概率为权的加权平均，也称**均值。**

针对密度函数则有：若 $\int_{-\infty}^{+\infty} |x| f(x) dx < +\infty$ 成立，则该变量的期望存在，且

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

本质上还是变量取值和概率乘积的求和。

若对于二维随机变量，也同理，只不过积分要积两次。

性质1： 若C是常数，则$E(C) = C$

性质2： $Y = aX + b$,则 $\mathbf{E}(\mathbf{Y}) = a\mathbf{E}(\mathbf{X}) + b$

性质3： $E(\mathbf{X} + \mathbf{Y}) = E(\boldsymbol{X}) + E(\mathbf{Y})$

$E(\mathbf{X} - \mathbf{Y}) = E(\boldsymbol{X}) - E(\mathbf{Y})$

$D(\mathbf{X} \pm \mathbf{Y}) = D(\boldsymbol{X}) + D(\mathbf{Y})$

性质4： $\boldsymbol{X}$和$\boldsymbol{Y}$不相关 则有 $E(\boldsymbol{XY}) = E(\boldsymbol{X})E(\boldsymbol{Y})$

且性质3和4可以扩展到n个随机变量。

随机变量的方差

2019年11月28日 19:45

方差

常用

$D(X) = E(X - E(X))^2 = E(X^2) - [E(X)]^2$

标准差

$\sigma(X) = \sqrt{D(X)}$

性质1：C是常数，$D(C) = 0$

性质2：$D(a\mathbf{X} + b) = a^2 D(\boldsymbol{X})$

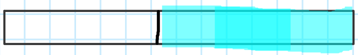
性质3：X和Y不相关，则$D(\mathbf{X} + \mathbf{Y}) = D(\mathbf{X}) + D(\mathbf{Y})$

因为在推导过程中用到了$E(\mathbf{X} + \mathbf{Y}) = E(\boldsymbol{X}) + E(\mathbf{Y})$，所以要相互独立。

接下来看一下[几种常见的分布的期望和方差]

切比雪夫不等式

设随机变量X的方差存在，则对任意常数$\varepsilon > 0$,有



ε

$P(|X - E(X)| < \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}$　　$P(|X - E(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}$

怎么记？记第二个式子。

如何理解：

括号里表示x与期望的差可以多大，然后再大频率也会不会和概率相差太远，即必然小于等于某个数。

记住第二个就可以推前一个式子。

希望通过方差来估计随机变量的取值与其数学期望之间的偏差大于某一正数的概率

该不等式给出了此概率的一个上界。

常见分布的期望和方差

2019年11月29日 13:51

1. **0-1 分布**
   $\mathrm{E(X) = p, D(x) = p(1-p)}$

2. **二项分布**
   $\mathrm{X} \sim B(n,p)$
   $\mathrm{E(X) = np, D(X) = np(1-p)}$

3. **泊松分布**
   $\mathrm{X} \sim P(\lambda)$
   $\mathrm{E(X)} = \lambda, D(X) = \lambda$

4. **几何分布**
   $\mathrm{E(X)} = \frac{1}{p}, D(X) = \frac{1-p}{p^2}$

5. **均匀分布**
   $\mathrm{X} \sim \boldsymbol{U}(a,b)$
   $\mathrm{E(X)} = \frac{a+b}{2}, D(X) = \frac{(b-a)^2}{12}$

6. **指数分布**
   $\mathrm{X} \sim E(\lambda)$
   $\mathrm{E(X)} = \frac{1}{\lambda}, D(X) = \frac{1}{\lambda^2}$　　　指数分布由其期望或方差所唯一确定

7. **正态分布**
   $\mathrm{X} \sim N(\mu, \sigma^2)$　　　正态分布的参数由其数学期望和方差唯一确定
   $\mathrm{E(X)} = \mu, \mathrm{D(X)} = \sigma^2$

协方差和相关系数

2019年11月28日 19:45

$Cov(\boldsymbol{X},\boldsymbol{Y})=E(\boldsymbol{X}-E(\boldsymbol{X}))E(\boldsymbol{Y}-E(\boldsymbol{Y}))=E(\boldsymbol{XY})-E(\boldsymbol{X})E(\boldsymbol{Y})$

称为XY的协方差。若协方差为0，则说明XY相互独立。
特别地，有$Cov(\boldsymbol{X},\boldsymbol{X})=D(\boldsymbol{X})$

$$\rho=\rho_{xy}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

记作XY的相关系数（此相关指线性相关）

性质1：对称性 $Cov(\boldsymbol{X},\boldsymbol{Y})=Cov(\boldsymbol{Y},\boldsymbol{X})$

性质2：对任意实数，a，b，有 $Cov(a\boldsymbol{X},b\boldsymbol{Y})=abCov(\boldsymbol{X},\boldsymbol{Y})$

性质3：结合律

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

计算多个随机变量线性组合的方差公式:

$\mathrm{D}(\mathrm{a}\boldsymbol{X}+\mathrm{b}\boldsymbol{Y}+\mathrm{c})=a^2D(\boldsymbol{X})+b^2D(\boldsymbol{Y})+2abCov(\boldsymbol{X},\boldsymbol{Y})$

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

随机变量的标准化

$$X^*=\frac{X-E(X)}{\sqrt{D(X)}}$$

称上述变换为**随机变量的标准化**，则$\mathrm{E}(X^*,Y^*)=\rho_{xy}$
即，标准化后的变量的协方差就是原变量的相关系数。

关于相关系数的性质

1. **基本性质**
   a. $|\rho_{xy}|\le 1$
   b. $|\rho_{xy}|=1$的充要条件是，存在常数a,b使得$P(\boldsymbol{Y}=a\boldsymbol{X}+b)=1$
   且$\rho_{xy}=1$ 时，$a>0$;$\rho_{xy}=-1$时，$a<0$。
2. **特殊性质**
   a. 若XY的相关系数为0（协方差为0），则称XY不（线性）相关。（可以有其他的相关）
   b. 下面的命题是等价的:
      i. $Cov(\boldsymbol{X},\boldsymbol{Y})=0$
      ii. X与Y不相关
      iii. $E(\boldsymbol{XY})-E(\boldsymbol{X})E(\boldsymbol{Y})=0$； $E(\boldsymbol{XY})=E(\boldsymbol{X})E(\boldsymbol{Y})$
      iv. $\mathrm{D}(\boldsymbol{X}+\boldsymbol{Y})=\mathrm{D}(\boldsymbol{X})+\mathrm{D}(\boldsymbol{Y})$

这里的不相关都是指不线性相关。

独立肯定不线性相关，（因为独立就是没有任何关系）
但是不线性相关却不一定独立。

独立是指没有任何关系。
但是不（线性）相关，却还可以存在其他相关，如平方等。

协方差矩阵与n维正态分布（无非是扩展罢了）但在多维正态分布中相当重要

其他数字特征

2019年11月28日 19:45

矩

原点矩

中心矩

分位数和中为数

变异系数

第五章 极限定理

2019年11月28日 19:45

大数定律

2019年11月28日 19:46

基本概念铺垫一下。
概率中的稳定性不同于微积分中的极限。
对于极限不在多说。
对于稳定性是指:
例如抛硬币的理想状态是正反面的概率均为$\frac{1}{2}$.
若取X = {正面朝上的次数} 的事件为一变量。
由于我们根据常识知道，X发生的概率为$\frac{1}{2}$。
其次把第 i 次试验记作X_i。即做了第 i 次实验。

现在忘掉概率的事实，假设我们不知道硬币正反的概率。
例如当你只做了一次实验时，即通过若干次独立实验确定了概率。可以抛很多次，（1,2,3，...）随你开心。这时你会得到一个概率，例如$\frac{1}{3}$
接下来你进行第二次实验，这次你抛了5次（随意的），得到概率为$\frac{2}{5}$(随机给一个而已)。
总而言之，你一共进行了i次实验，每次实验抛硬币的次数都由你自己随意决定。
当你实验次数够多时，所有概率的算术平均值就会趋近于真正的概率。
当然我们知道概率肯定是$\frac{1}{2}$，所以就能想象出，当你实验的次数足够多时，便更能得出正确的概率结果。
所以就需要i足够大。

这时我们还能得出
$$\lim_{i\to+\infty} P(|X_i - X| \geq \varepsilon) = 0$$
称为X_i**依照概率收敛于*X***。 $X_i - X \xrightarrow{P} 0$
这里的***X***我们知道是 $\frac{1}{2}$。即不管你做了多少次实验，所得结果于$\frac{1}{2}$的差可以取到任意小甚至是 0 。
而且当i越大时，就越精准，越有说服力。当i很大时，我们就能有很大的把握保证X_i和X的偏差很小，两者会很接近。

通俗的说就是你进行足够多的实验后，粗略地得出概率会是$\frac{1}{2}$，然后你又看到有人要抛硬币，但是你根据自己的结果已经可以有很大的把握保证，他得出的结果和你的偏差不会很大。

但是这与极限不同的是，极限是越接近于无穷，你越能确定它的取值。
就如 y = $\frac{1}{x}$，当 x 充分大时， y 就越接近于 0 。
但是对于概率，虽然你知道正面向上的概率是$\frac{1}{2}$，但即使抛五次也有可能全是正面，这种情况依然存在，但是在你抛的次数足够多的时候，出现这种偏差的概率就会越小，即越接近于$\frac{1}{2}$。

大数定律

$\{X_n: n = 1,2,\ldots\}$是一列随机变量序列，如果存在一列实数$\{a_n\}$使得
$$\frac{1}{\mathrm{n}}\sum_{i=1}^{n} X_i - a_n \xrightarrow{P} 0$$
成立，则称改序列服从**大数定律**。

对变量限制小，但是成立条件较苛刻。

切比雪夫大数定律

设$\{X_n: n = 1,2,\ldots\}$是一列**相互独立**的随机变量序列，若存在常数C，均有$D(X_i) \leq C$，成立，则对于任意$\epsilon > 0$，有
$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| \geq \varepsilon\right) = 0$$
利用算术平均值来接近实际真值，所谓的多测几次取平均值减小误差。

设$\{X_n: n = 1,2,\ldots\}$是一列**独立同分布**的随机变量序列，且$E(X_k) = \mu, D(X_k) = \sigma^2$，由切比雪夫大数定理，有
$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right) = 0$$
前一项是n个观察值的平均值，而u是被测物的真实值，因此该定理就是说明：当试验次数趋于无穷时，实际测量值的算术平均值会依概率收敛于真实值。即，

设$\{X_n: n = 1,2,\ldots\}$是一列**独立同分布**的随机变量序列，且$E(X_k) = \mu, D(X_k) = \sigma^2$，由切比雪夫大数定理，有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right) = 0$$

前一项是n个观察值的平均值，而u是被测物的真实值，因此该定理就是说明：当试验次数趋于无穷时，实际测量值的算术平均值会依概率收敛于真实值。即，多测几次取算术平均值来作为被测量值的近似值。

即

$$\frac{1}{\mathrm{n}}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

伯努利大数定律

设X_n是n次独立实验中事件A发生的次数，p是事件A发生的概率，则对于任意$\epsilon > 0$，有

用频率来接近概率 抛硬币例子

针对二项分布。

$$\lim_{n\to\infty} P\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) = 0$$

$$\frac{X_n}{n} \xrightarrow{P} p$$

辛钦大数定律

设$\{X_n: n = 1,2,\ldots\}$是一列**独立同分布**的随机变量序列，且$E(X_k) = \mu$，则对于任意$\epsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right) = 0$$

就是特殊情况下的切比雪夫大数定律

对变量限制大，但成立条件较弱。

$$\frac{1}{\mathrm{n}}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

伯努利大数定律是300年前瑞士数学家伯努利潜心研究20年证明出来的，是人类历史上第一个严格证明的大数定律。它是辛钦大数定律的特殊情况，不过由于它有一定的历史意义并且二项分布的大数定律在日常生活中最为常见，所以编教材的人喜欢把这个大数定律单独列出来。

切比雪夫大数定律和辛钦大数定律针对的是两种不同的情况，谁也不是谁的特例。切比雪夫大数定律说的是一列独立变量（可以不同分布）的均值收敛到一个常数，但前提是每个变量的期望和方差均存在且有限，并且满足方差的平均值是样本数n的高阶无穷小这一额外条件。辛钦大数定律是说一列独立同分布的随机变量的均值收敛到一个常数，条件是分布的绝对期望存在且有限就够了。

对两个大数定律做一总结，就是切比雪夫大数定律不要求随机变量有相同分布但是成立的条件更加严格，辛钦大数定律要求同分布不过是在比较弱的条件下就成立。

中心极限定律

2019年11月28日 19:46

确定在什么条件下的打量的随机变量之和的分布可以用正态分布来近似

林德伯格-列维中心极限定理（针对一系列事件，有多个样本 。一般变量设为第i天...，第i次实验...，第i样...）

设$\{X_n\}$是**一独立同分布**的随机变量序列，且$\mathrm{E}(X_k) = \mu, D(X_k) = \sigma^2 < +\infty$，则对任意x，有

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \le x\right) = \phi(x)$$

仔细观察括号里的式子，类似于随机变量的标准化。x的和服从$(n\mu, n\sigma^2)$

其中$\phi(x)$是标准正态的分布函数

棣莫佛-拉普拉斯中心极限定理（针对一个事件）

设Y_n是n次独立实验中事件A发生的次数，又在每次事件中A发生的概率为p，则对任意给定实数x，有

$$\lim_{n\to\infty} P\left(\frac{Y_n - np}{\sqrt{npq}} \le x\right) = \phi(x) \qquad q = 1 - p$$

事实上，$Y_n \sim B(n,p)$，但是$\frac{Y_n - np}{\sqrt{npq}}$近似服从$\mathrm{N}(0,1)$，或者说$Y_n \sim N(np, npq)$,所以就把二项分布的问题转为了正态分布计算的问题。

第六章 抽样分布

2019年12月6日 20:07

总体与样本

2019年12月6日　　20:07

要测量的全体是总体。
从中抽取一部分为样本。
且两者具有相同的分布。

统计量与抽样分布

2019年12月6日 20:07

统计量

统计量应为一直切实的数据，不含未知参数。

样本平均值

就是平均值公式

样本方差

与方差公式有略微区别

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

概率是频率趋于无穷的表现。

期望是均值趋于无穷的表现。

样本标准差

样本方差的算术平方根

样本k阶矩

$$A_k = \frac{1}{n}\sum_{i=1}^{n}X_i^k$$

样本k阶中心矩

$$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$$

当样本容量趋于无穷时，样本k阶矩依概率收敛于总体k阶矩。（第七章）

定义： 把样本数据从小到大排列，取第k个数据，称其为**第k为顺序统计量。**

特别地，第一位又称**最小位统计量**。第n位又称**最大位顺序统计量。**

根据第三章的最大值和最小值分布函数，有

最大位统计量

$$F_M(x) = F_{X_1}(x)\dots F_{X_2}(x) = [F(x)]^n$$

注意此时，这里的x同分布，所以就有相同的分布函数。

最小位统计量

$$F_N(x) = 1-[1-F_{X_1}(x)]\dots[1-F_{X_2}(x)] = 1-[1-F(x)]^n$$

定义6.2.4：经验分布函数

设$X_1, X_2, \dots, X_n$是来自总体X的样本，对于任意的实数x，用S(x)表示样本中不大于x的随机变量的个数，则S(x)表示事件{X≤x}出现的频数，而它出现的频率

$$P_X(X \le x) = F_X(x) = \frac{1}{n}S(x) = \begin{cases} 0, x < X_{(1)} \\ \frac{k}{n}, X_{(k)} \le x \le X_{(k+1)} \\ 1, x \ge X_{(n)} \end{cases}$$

例：

从一批标准重量为5009的罐头中，随机抽取8听，测得误差如下(单位；g)：8，—4，6，—7，—2，1，0，1，求经验分布函数，并作出图形。

解：将样本值按大小顺序排列为：—7<—4<—2<0<1=1<6<8

则其样本分布函数为：

$$F_8^*(x)=\begin{cases}0 & x<-7\text{时}\\ \frac{1}{8} & -7\leqslant x<-4\\ \frac{2}{8} & -4\leqslant x<-2\\ \frac{3}{8} & -2\leqslant x<0\\ \frac{4}{8} & 0\leqslant x<1\\ \frac{6}{8} & 1\leqslant x<6\\ \frac{7}{8} & 6\leqslant x<8\\ 1 & x\geqslant 8\end{cases}$$

定理6.2.2:对于任意实数x，当n趋于无穷时，经验分布函数以概率1趋于总体的分布函数。

所以当样本取得够大时，得出的分布函数就可以作为估计运用于整体。

分布

在总体X服从正态分布时，求一些统计量的精确分布。（自由度又称参数）

χ^2分布

定义6.2.5：设随机变量$X_1,\dots,X_n$相互独立且均服从标准正态分布N(0,1)，则称随机变量

$$\boldsymbol{\chi^2 = X_1^2+\cdots+X_n^2}$$

所服从的分布时**自由度为n**的χ^2分布，记为$\chi^2\sim\chi^2(n)$

可加性：设$\chi_1^2\sim\chi_1^2(n_1),\chi_2^2\sim\chi_2^2(n_2)$，且相互独立，则有$\chi_1^2+\chi_2^2\sim\chi_1^2(n_1+n_2)$

和正态独立分布的可加性稍作区别看待。但是异曲同工。

$$\mathbf{E}(\chi^2)=\mathbf{n}$$

$$\mathbf{D}(\chi^2)=\mathbf{2n}$$

t分布

定义6.2.6：设$X\sim N(0,1),Y\sim\chi^2(n)$,并且$X$于$Y$独立，则称随机变量

$$t=\frac{X}{\sqrt{Y/n}}$$

所服从的分布是**自由度为n的t分布**，记为$t\sim t(n)$

t(n)分布的概率密度函数是偶函数。因此t分布时对称分布。

当n>30时，t分布就很接近标准正态分布了。

F分布

设$X\sim\chi^2(n_1),Y\sim\chi^2(n_2)$且X于Y相互独立，则称随机变量

$$F=\frac{X/n_1}{Y/n_2}$$

所服从的分布是**自由度为(n_1,n_2)的分布**，记为$F\sim F(n_1,n_2)$。

显然若$F \sim F(n_1, n_2)$，则

$$\frac{1}{F} \sim F(n_1, n_2)$$

正态总体的抽样分布

2019年12月6日 20:07

设$(X_1,\dots,X_n)$是取自正态总体X~N(μ,σ^2)的样本，$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 为其样本均值，$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 为其样本方差。

定理6.3.1： 若X~N(μ,σ^2),则 $\bar{X}$~N(μ,$\frac{\sigma^2}{n}$)。

推论：

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

定理6.3.2： 若X~N(μ,σ^2),则

(1) $\bar{X}$与S^2相互独立;

(2) $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

定理6.3.3： 若X~N(μ,σ^2),则

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

由上面的定理和推论可得。

定理6.3.4： 设$(X_1,\dots,X_n)$是取自正态总体X~N(μ_1,σ_1^2)的样本，$(Y_1,\dots,Y_n)$是取自正态总体X~N(μ_2,σ_2^2)的样本,且相互独立。

(1) $\mathrm{F}=\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$

(2) 当$\sigma_1^2=\sigma_2^2$时，有

$$\mathrm{T}=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

其中,

$$S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

注意: 由取自正态总体N(μ,σ^2)的样本$X_1,X_2\dots X_n$所构造的两个函数 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$ 与 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 之间的区别，它们分别服从自由度为n与n-1的χ^2分布。

第七章 参数估计

2019年12月18日　21:14

点估计

2019年12月18日　　21:18

7.1.1矩法

说得通俗点就是

一阶原点矩就是样本的均值,二阶中心矩就是样本的方差.

矩估计就是用样本的均值估计总体的期望,用样本的统计量（方差）估计总体的方差

有n个未知量就要n个矩方程，就要到 X的n次 的期望。

7.1.2最大似然法

对于离散型随机变量，累乘概率。

对于连续性随机变量，累乘概率密度函数。 都要趋于0，若不等于0，则利用单调性，找到最接近的。

估计量的评判标准

2019年12月18日 21:18

对于统一参数，由不同的方法可以得到不同的估计量，如何才能找到最接近实际值的估计量，则需要用到下面三个性质。

7.2.1 无偏性

估计量与实际值波动性小，即更稳定。

就要求

$E(\hat{\theta})=\theta$ 无偏估计量

$\lim_{n\to\infty} E(\hat{\theta})=\theta$ 渐进无偏估计量

7.2.2 有效性

有效性则表示波动幅度越小越好。

即对于不同的无偏估计量，方差越小越好。

至于方差能小到什么程度，是否有下界，则由C-R不等式确定。

7.2.3 一致性

设$\hat{\theta}(X_1,X_2\ldots X_n)$是$\theta$的估计量，若对任意给定的正数ε，有

$$\lim_{n\to\infty} P\{|\hat{\theta}(X_1,X_2\ldots X_n)-\theta|\geq\varepsilon\}=0$$

对一切θ的估计量都成立，当n趋于无穷时依概率收敛于θ，则称该估计量是**一致估计量**。

区间估计

2019年12月18日 21:19

由于点估计出来的估计量也只是一个估计值。由于样本的随机值，估计量也只在一个随机区间内。
为了确定该范围内包含真值的概率，就有了区间估计。

定义7.3.1： 设总体X的分布函数为$\mathrm{F}(x;\theta)$,θ是未知参数，$X_1,X_2\dots X_n$是来自X的样本。α是给定值(0<α<1)，若有两个统计量$\underline{\theta}=\underline{\theta}(X_1,X_2\dots X_n)$,$\overline{\theta}=\overline{\theta}(X_1,X_2\dots X_n)$满足

$$\mathrm{P}\left\{\underline{\theta}<\theta<\overline{\theta}\right\}=1-\alpha$$

则称随机区间$(\underline{\theta},\overline{\theta})$是θ的置信度为1-α的**置信区间**。分别称其为**置信上限**和**置信下限。**

正态总体均值与方差的区间估计

2019年12月18日 21:19

单侧置信区间

2019年12月18日　　21:19

第八章 假设检验

2019年12月18日 21:19

假设检验的基本概念

2019年12月18日 21:19

单个正态总体参数的假设检验

2019年12月18日　21:19

分布拟合检验

2019年12月18日　21:20