概率统计公式-1

- 第二章
  - 分布函数F(x)
    - $F(x) = P(X<=x) = ∫(-∞,x) f(t)dt$
      - $P(a<x<b) = F(b)-F(a) = ∫(a,b)f(x)dx$
    - 性质
      - 单调不减
      - 负无穷极限=0，正无穷极限=1
      - 右连续
      - 任意一点的概率= F(x)-F(x0-0)
  - 离散型随机变量概率分布
    - 分布律
      - 表格，矩阵
      - 如何用分布律算分布函数?
        - 查表，加在一起
      - 如何计算分布律?
        - S1: 计算X所有可能取值
        - S2: 计算每个X对应的概率
        - S3: 写成表格/矩阵
    - 常见的离散分布
      - 几何分布
      - 超几何分布
      - 二项分布
      - 泊松分布
  - 连续型随机变量
    - 概率密度
      - $F(x) = ∫f(t)dt$
        - $F'(x) = f(x)$
          - $P(a<x<b) = F(b)-F(a) = ∫(a,b)f(x)dx$
      - 性质
        - $f(x)>=0$
        - $∫f(x)dx = 1$
          - 算c

- P(x)=0
- 三种常见的连续分布
  - 均匀分布
  - 指数分布
    - 无记忆性
  - 伽玛分布
    - γ(a) = ∫(0,+∞) x^(a-1) e^(-x) dx
    - γ(a+1) = aγ(a)
    - γ(1)=1, γ(1/2)=√π
    - n∈Z+, γ(n)=(n-1)!
- **随机变量函数的分布（压轴）**
  - 题型
    - 设X分布已知，Y=g(X),求Y分布
  - 离散X
    - 列表X和P
    - 中间加一行Y
    - 合并相同的，最终得到Y~P
  - 连续X
    - 情况一：Y=g(X)在X取值区间上单调
      - 直接解不等式
    - 情况二：Y=g(X)在X取值区间不单调且对称
      - 解第一个不等式的时候画图，第二个不等式的时候增加辅助线Y=y
    - 情况三：Y=g(X)在X取值区间不单调且不对称
    - S1: 计算Y的取值区间
    - S2: 计算Y的分布函数
      - FY(y)=P(Y<=y)=P(代入<=y)
      - 解出括号内x的取值范围
      - 再转化为密度函数积分
        - 可以保留变上限积分，也可以算出来
    - S3: 计算Y的密度函数
      - fY(y)=F'Y(y)
- 第三章
  - 联合分布函数F
    - 定义：F(x,y)=P{(X<=x)∩(Y<=y)} = P(X<=x, Y<=y)
    - 性质

- F关于x，y分别单调不减
- 有界性
  - [0,1]
  - F(x,+∞)=F(-∞,y)=0
  - F(-∞,-∞)=0, F(+∞,+∞)=1
- 右连续
- **非负性**

$$F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)\geq 0$$

- 二维离散型随机变量及其概率分布
  - 联合分布律
    - i行j列的表格
    - 性质
      - 非负
      - 归一
  - **如何计算离散型联合分布律**
    - 同上
- 二维连续型随机变量
  - 定义

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

  - 性质
    - 归一性
      - **求c**
    - 非负
    - P{(X,Y)∈G}=∬f(x,y)dxdy
    -

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\partial y}$$

  - **如何计算连续型联合分布律**
    - S1: 画出联合F的取值区间
    - S2: 找最优积分方向，改写重积分

      先积谁就用平行于谁的箭头穿过积分区域
  - **常见分布：二维均匀分布**

$$f(x,y)=\begin{cases}\dfrac{1}{m(G)}, & (x,y)\in G\\ 0, & 其他\end{cases}$$

- 边缘分布函数&边缘概率密度
  - **如何联合推边缘?**
    - 离散
      - 算分布律：给联合的二维中心表，求和得到周围一圈
    - 连续
      - 把+∞代入 FX(x)= F(x,+∞)
      - fX(x)=∫fXY(x,y)dy
        - 积分上下界为边界表达式
        - 取值范围为x的取值区间最值点
  - **如何边缘推联合?**
    - 离散
      - S1: 需要自己推断独立性
      - S2: 画成周围一圈，算出中心那块儿（分布律）
    - 连续型的题会给独立性
      - X与Y独立 等价于 F(x,y)=FX(x)*FY(y)
      - X与Y独立 等价于 f(x,y)=fX(x)*fY(y)
- 条件分布&条件密度
  - F

$$F_{X|Y}(x\mid y)=P(X\le x\mid Y=y),x\in R$$

  - 求离散的分布律
    - S1: 画联合XY二维表
    - S2: 写公式，查表圈分子分母
    - S3: 化简计算
  - 求连续的密度函数
    - 条件分布函数

$$F_{X|Y}(x\mid y)=\frac{\int_{-\infty}^{x} f(u,y)du}{f_Y(y)}$$

    - 求条件密度函数：条件密度 = 联合密度 / 边缘密度

$$f_{X|Y}(x\mid y)=\frac{\partial F_{X|Y}(x\mid y)}{\partial x}=\frac{f(x,y)}{f_Y(y)}\quad (f_Y(y)>0)$$

      - S1: 画联合分布函数的取值区间
      - S2: 求边缘密度函数

- S3: 求边缘密度的自变量取值区间【为最值点】，作为大前提
- S4: 求条件密度函数的自变量取值区间
  - 注意永远是前一个，X|Y就是求X的
  - 【为取值区域的边界表达式】
- 标准过程：
  - 当（边缘密度函数自变量取值区间）时
  - （条件密度函数表达式，附其自变量取值区间）
  - 【如果还需求P】
  - 当（边缘密度函数自变量=多少）时
  - （代入算出的条件密度函数表达式，记得附其自变量取值区间）
  - P(|)=∫====
- **难点：找准取值区间**

- **二维随机变量函数的分布**

  已知(X,Y)的分布，求Z分布
  - 离散
    - S1: 画一张大表，第一行(X,Y)所有可能取值，第二行P第三行Z
    - S2: 合并相同的Z，得到Z的分布
  - **连续**
    - S1: 计算Z的取值区间
    - S2: 计算Z的F
      - FZ(z)=P(Z<=z) = P(代入<=z)=P(X,Y的取值区域)
      - 判断形状是否变化，决定是否分类讨论
      - 改写成积分
- 多维随机变量
  - N维随机变量的分布
  - 分布的可加性
    - X1...Xn同分布，若X1+...+Xn也同其分布，则有可加性
      - 二项分布有可加性

        $$X_1 + X_2 + \cdots + X_n \sim B(\sum_{i=1}^{n} n_i, p)$$
      - 01分布有可加性

        $$X_1 + X_2 + \cdots + X_n \sim B(n, p)$$
      - 泊松分布有可加性

$$Z = X + Y \sim P(\lambda_1 + \lambda_2)$$

- γ分布具有可加性

$$X_1 + X_2 + \cdots + X_n \sim \Gamma(\sum_{i=1}^{n} \alpha_i, \beta)$$

- **最大值最小值函数的分布**
  - 最大值最小值定义

    $$M = \max\{X_1, X_2, \cdots, X_n\}, N = \min\{X_1, X_2, \cdots, X_n\}$$

  -

    $$若 X_1, X_2, \cdots, X_n 独立同分布于 F(x),$$
    $$F_M(x) = F^n(x)$$
    $$F_N(x) = 1 - [1 - F(x)]^n$$

- 第四章 随机变量的数字特征
  - 期望
    - 离散型的期望

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

      - ∑数值*概率
    - 连续型的期望

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

    - 常见分布的期望
      - p
      - np
      - λ
      - (a+b)/2
      - 1/λ
      - α/β
    - 随机变量函数的期望
      - 一维

        $$E(Y) = E[g(X)] = \begin{cases} \sum_{k=1}^{\infty} g(x_k) p_k, & X离散型 \\ \int_{-\infty}^{\infty} g(x) f(x) dx, & X连续型 \end{cases}$$

        - 不需要知道Y的分布，也可以求E

- 二维

  (1) (X,Y)是离散型的，联合分布律为$p_{ij}, i,j=1,2,\cdots$，

  $$E(Z)=E(g(X,Y))=\sum_i\sum_j g(x_i,y_j)p_{ij};$$

  (2) (X,Y)是连续型的，联合密度函数为$f(x,y)$

  $$E(Z)=E(g(X,Y))=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy.$$

  - 离散的就直接带进去算就行了

- 期望的性质
  - C的期望E(C)

    $$E(C)=C$$

  - 期望内的常数可以提出来

    $$(2)\ E(CX)=CE(X)$$

  - 和的期望=期望的和

    $$(3)\ E(X+Y)=E(X)+E(Y)$$

  - 线性性质

    $$E(\sum_{i=1}^{n} C_i X_i+b)=\sum_{i=1}^{n} C_i E(X_i)+b$$

  - 独立则乘积（反过来不一定成立）

    若X,Y相互独立，则 $E(XY)=E(X)E(Y)$

- 方差
  - 方差的定义

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

    - 均方差/标准差

      而$\sqrt{D(X)}$称为X的均方差或标准差

  - 方差的算法及简便算法

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

  - 方差的性质
    - 常数的方差为0

      $$D(C)=0$$

    - 不具有线性性质

      $$D(aX+b)=a^2D(X).$$

- 独立时加减可拆

  $$D(X \pm Y) = D(X) + D(Y).$$

- 方差=0的充要条件

  (5) $D(X) = 0$的充要条件是存在常数C，使得$P(X = C) = 1$，且$C = E(X)$。

- 常见分布的方差
  - p(1-p)
  - np(1-p)
  - λ
  - (b-a)^2/12
  - 1/λ^2
  - α/β^2
- 随机变量标准化、变异系数、中心矩、原点矩
  - 标准化

    $$X^* = \frac{X - E(X)}{\sqrt{DX}} \boxed{= \frac{X - c}{d}}$$

    线性变换

  - 变异系数

    $$C_v = \frac{\sqrt{D(X)}}{|E(X)|}$$

  - 原点矩和中心矩

    $$m_k = E(X^k),$$

    称m_k为X的k阶原点矩。

    $$\mu_k = E[X - E(X)]^k,$$

    称μ_k为X的k阶中心矩。

    - 中心矩μ计算方法
      - $$D(X) = \mu_2$$
      - $$\mu_3 = m_3 - 3m_2m_1 + 2m_1^3$$
- 协方差和相关系数
  - 协方差

      - 定义
        - Cov(X,Y)=E{[X-E(X)][Y-E(Y)]}
        - Cov(X,X) = D(X)
      - 计算方法
        - Cov(X,Y) = E(XY)-E(X)E(Y)
      - 性质
        - 和常数：Cov(X,C)=0
        - 数量积：Cov(aX,bY)=abCov(X,Y)
        - 可加性：Cov(X+Y,Z)=Cov(X,Z)+Cov(Y,Z)
      - (X,Y)均值向量为(E(X),E(Y))
      - (X,Y)的协方差阵为D(X) Cov(X,Y) // D(Y) Cocv(Y,X)
    - 相关系数
      - 性质
      - 计算方法
        - R = Cov(X,Y)/√DX √DY
- 第一章
  - 独立性
    - 期望：独立则乘积，反之不成立

      若X,Y相互独立，则 $E(XY)=E(X)E(Y)$
- 第五单元 正态分布
  - 定义和密度函数

正态分布	密度函数
$X\sim N(0,1)$	$\varphi(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\quad x\in R$
$X\sim N(\mu,\sigma^2)$	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad x\in R$

  - 性质

$X\sim N(0,1)$	(1) $\Phi(0)=\frac{1}{2}$; (2) $\Phi(-x)=1-\Phi(x)$. (3) $P(a<X\le b)=\Phi(b)-\Phi(a)$
$X\sim N(\mu,\sigma^2)$	(1) $F(x)=\Phi\left(\frac{x-\mu}{\sigma}\right)$ (2) $P(a<X\le b)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$

- 线性性质

  设随机变量 $X\sim N(\mu,\sigma^2)$，当 $b\neq 0$ 时，有

  $$Y=a+bX\sim N(a+b\mu,b^2\sigma^2).$$

- 第六单元 极限定理
  - 切比雪夫不等式
  - 独立同分布大数律
  - 林德伯格-列维中心极限定理
  - 棣莫弗-拉普拉斯中心极限定理
- 第七单元 数理统计基础
  - 三大分布
    - χ2分布
      - 定义：X~N(0,1)，∑(Xi)2 ~ χ2(n)
      - γ(n/2,1/2)
      - 可加性
      - Eχ2=n, Dχ2=2n
    - t分布
      - 定义：X~N(0,1),Y~χ2(n), X/√(Y/n) ~t(n)
      - n>1， EX = 1
      - n充分大， X/√(Y/n)~N(0,1)
    - F分布
      - X~χ2(n),Y~χ2(m)， 则(X/n)/(Y/m) ~F(n,m)
      - 1/F ~ F(m,n)
      - X~t(n),则X^2~F(1,n)
  - *分位点*
  - 统计量
  - 抽样分布定理

    - 正态总体
    - 任何总体
- 第八单元 估计
  - 点估计
    - 矩估计
      - 一个未知参数估计
      - 两个未知参数估计
    - *极大似然估计*
      - 方程有解
      - 方程无解
  - 估计量的评选标准
    - 无偏性标准
      - EA = B

        > A是B的无偏估计
      - 没有函数传递性
      - 一个参数的无偏估计可以有多个
      - 偏差EA-B
      - 渐进无偏估计量 n->∞， EA-B->0
    - 有效性标准
      - EA = C且EB=C
      - DA <= DB， A更有效
    - 一致性标准
      - A是B的一致估计量，则A依概率收敛到B
    - 均方误差标准
      - 计算

        $$M(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

        - 换成D(θ尖-θ)+E^2(θ尖-θ)=D(θ尖)-[E(θ尖)-θ]^2
          - 也就是算θ尖的D和E
      - 比较有效

        $$M(\hat{\theta}_1) \leq M(\hat{\theta}_2)$$

        则称 $\hat{\theta}_1$ 在均方误差下比 $\hat{\theta}_2$ 更有效。

- 区间估计
  - 一个正态总体下的区间估计

    假设X服从$N(\mu,\sigma^2)$，$X_1,X_2,\cdots,X_n$来自总体X的一个样本。

    1. 已知$\sigma^2=\sigma_0^2$，求总体均值μ的置信度为1-α的置信区间

    $$\left(\bar{X}-u_{1-\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}},\bar{X}+u_{1-\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}}\right).$$

    2. σ^2未知，求总体均值μ的置信度为1-α的置信区间

    $$\left(\bar{X}-t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}},\bar{X}+t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right).$$

    3. μ未知时，求总体方差σ^2的置信区间

    $$\left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)},\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right)$$

  - 两个正态总体下的区间估计

    设有两个正态总体$X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)$。$X_1,X_2,\cdots,X_{n_1}$和$Y_1,Y_2,\cdots,Y_{n_2}$是分别来自$X$和$Y$的两个独立样本。

    1. σ_1^2和σ_2^2都已知，求总体均值差$\mu_1-\mu_2$的置信度为$1-\alpha$的置信区间

    $$\left(\bar{X}-\bar{Y}-u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\bar{X}-\bar{Y}+u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

    2. σ_1^2和σ_2^2都未知，但$\sigma_1^2=\sigma_2^2=\sigma^2$，求$\mu_1-\mu_2$的置信度为$1-\alpha$的置信区间

    $$\left(\bar{X}-\bar{Y}-t_{1-\frac{\alpha}{2}}S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ \bar{X}-\bar{Y}+t_{1-\frac{\alpha}{2}}S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

    3. $\mu_1,\mu_2,\sigma_1^2,\sigma_2^2$都未知，求总体方差比$\frac{\sigma_1^2}{\sigma_2^2}$的置信度为$1-\alpha$的置信区间

    $$\left(\frac{S_1^2}{S_2^2F_{1-\frac{\alpha}{2}}},\frac{S_1^2}{S_2^2F_{\frac{\alpha}{2}}}\right)$$

- 第九单元
  - 一个正态总体下，假设检验拒绝域

    1/2. σ^2已知/未知时，μ的检验

    给定显著性水平α，在σ^2已知和σ^2未知两种情况下关于H_0的拒绝域如下表所示

H_0	H_1	σ^2已知	σ^2未知
$\mu=\mu_0$	$\mu\neq\mu_0$	$\lvert U\rvert>u_{1-\frac{\alpha}{2}}$	$\lvert t\rvert>t_{1-\frac{\alpha}{2}}(n-1)$
$\mu=\mu_0$	$\mu>\mu_0$	$U>u_{1-\alpha}$	$t>t_{1-\alpha}(n-1)$
$\mu=\mu_0$	$\mu<\mu_0$	$U<-u_{1-\alpha}$	$t<-t_{1-\alpha}(n-1)$

    注意：在各种备择假设下，拒绝域的方向与备择假设的方向一致。

3. μ未知时，σ^2的检验

当原假设为$H_0: \sigma^2 = \sigma_0^2$，且$H_0$成立时，由抽样分布定理7.4，可知检验统计量

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$$

对给定的显著性水平α，下表给出了各种备择假设下检验的拒绝域。这种检验叫做χ^2检验。

H_0	H_1	显著性水平α下关于H_0的拒绝域
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi^2_{\frac{\alpha}{2}}(n-1)$或$\chi^2 > \chi^2_{1-\frac{\alpha}{2}}(n-1)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{1-\alpha}(n-1)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{\alpha}(n-1)$

- 两个正态总体下，假设检验拒绝域

1/2. σ_1^2, σ_2^2已知/σ_1^2, σ_2^2未知，但$\sigma_1^2=\sigma_2^2=\sigma^2$时，$\mu_1=\mu_2$的检验

在显著性水平α的条件下，H_0的拒绝域

H_0	H_1	σ_1^2, σ_2^2已知	σ_1^2, σ_2^2未知，$\sigma_1^2=\sigma_2^2$
$\mu_1=\mu_2$	$\mu_1 \neq \mu_2$	$\lvert U\rvert > u_{1-\alpha/2}$	$\lvert T\rvert > t_{1-\alpha/2}(n_1+n_2-2)$
$\mu_1=\mu_2$	$\mu_1 > \mu_2$	$U > u_{1-\alpha}$	$T > t_{1-\alpha}(n_1+n_2-2)$
$\mu_1=\mu_2$	$\mu_1 < \mu_2$	$U < -u_{1-\alpha}$	$T < -t_{1-\alpha}(n_1+n_2-2)$

注意：在各种备择假设下，拒绝域的方向与备择假设的方向一致。

3. μ_1, μ_2未知，求总体方差$\sigma_1^2 = \sigma_2^2$的假设检验

当原假设为H_0：$\sigma_1^2 = \sigma_2^2$成立时，由抽样分布定理可以得到此时的检验统计量

$$F = \frac{S_1^2}{S_2^2}\text{服从}F(n_1-1, n_2-1).$$

因此，这种假设检验叫F检验。

在显著性水平α的条件下，两正态总体方差检验H_0的拒绝域

H_0	H_1	H_0的拒绝域
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$F < F_{\alpha/2}(n_1-1, n_2-1)$或$F > F_{1-\alpha/2}(n_1-1, n_2-1)$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$F > F_{1-\alpha}(n_1-1, n_2-1)$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$F < F_{\alpha}(n_1-1, n_2-1)$