概率论

统计量与抽样分布

正态分布的一些性质

两个独立的正态分布，和也是正态分布。

正态分布的k阶原点矩

$$X \sim N(0,1), E(X^k) = (k-1)!!, k\text{是偶数}; E(X^k) = 0, k\text{是奇数}$$

正态总体

χ^2分布

$\chi_n^2 = \sum\limits_{i=1}^{n} X_i^2$

性质:

1. $\chi_1^2 \sim \chi^2(n_1), \chi_2^2 \sim \chi^2(n_2)$且$\chi_1^2, \chi_2^2$相互独立，则有$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$
2. $\chi^2 \sim \chi^2(n) \Rightarrow E(\chi^2) = n, D(\chi^2) = 2n$

t分布

$T = \dfrac{X}{\sqrt{Y/n}}, X \sim N(0,1), Y \sim \chi^2(n)$,且X和Y相互独立。

关于y轴对称。

F分布

$F = \dfrac{U/n_1}{V/n_2}, U \sim \chi^2(n_1), V \sim \chi^2(n_2)$且U和V相互独立。$F \sim F(n_1, n_2)$

性质:

1. $F \sim F(n_1, n_2) \Rightarrow \dfrac{1}{F} \sim F(n_2, n_1)$
2. $T \sim t(n) \Rightarrow T^2 \sim F(1, n)$

上分位点

$P(X > \lambda_\alpha) = \alpha$,$\lambda_\alpha$为X的$\alpha$分位点。

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

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

$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)}$$

正态总体的样本均值与样本方差的分布

设$X_1, X_2, ..., X_n$是来自正态总体$N(\mu, \sigma^2)$的一个样本，则

$$\bar{X} \sim N(\mu.\frac{\sigma^2}{n})$$

$$\frac{nS_n^2}{\sigma^2} \sim \chi^2(n-1) \quad or \quad \frac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\bar{X}与S^2相互独立$$

$$设X_1, X_2, ..., X_n是来自正态总体N(\mu, \sigma^2)的一个样本。则T = \frac{(\bar{X}-\mu)}{S_{n-1}/\sqrt{n}} \sim t(n-1)$$

$$X_1, .., X_{n_1}是来自正态总体N(\mu_1, \sigma_1^2)的一个样本, Y_1, .., Y_{n_2}是来自正态总体N(\mu_2, \sigma_2^2)的一个样本，且两样本相互独立，$$

$$记S_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$$

$$记S_2^2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$$

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

参数估计

矩估计

以样本矩作为总体矩的估计从而得到参数的估计量

有几个参数就求几阶原点矩，然后得到方程组求解。

估计值在参数上面加一个$\hat{\lambda}$

注意：方差和期望之间的转换方式，以及样本方差S_{n-1}^2和S_n^2的不同，这里用的是后者。

无论总体X服从何种分布，总体均值$EX = \mu$, 总体方差$DX = \sigma^2$作为未知参数，其矩估计量一定是样本均值和样本方差，即

$$\hat{\mu} = \overline{X}, \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = S_n^2$$

相关系数的矩估计：

$$\rho_{XY} = \frac{cov(X,Y)}{\sqrt{D(X)D(Y)}} = \frac{E((X-E(X))(Y-E(Y)))}{\sqrt{(E((X-EX)^2)E((Y-E(Y))^2)}}$$

然后用$\bar{X}$和S_n^2替换.

矩估计特殊情况

一阶不行时求二阶。

极大似然估计

选择出现样本情况概率最高的参数取值。

求出最大似然函数，对每个参数求偏导可得。

连续性随机变量，将概率密度相乘即可。

离散型随机变量将分布律相乘。

极大似然估计的不变性

设$\hat{\theta}$是θ的极大似然估计，$u(u(\theta))$是θ的函数，且有单值反函数：$\theta=\theta(u)$,则$\hat{u}=u(\hat{\theta})$是$u(\theta)$的极大似然估计。

$\hat{\theta}$是θ的极大似然估计，则$u(\hat{\theta})$是$u(\theta)$的极大似然估计

如果极大似然方程组无解，可以直接考虑极大似然函数，使其最大，求得其最大时参数的取值（例如均匀分布的极大似然估计）

估计量的评选标准

无偏性

$E(\hat{\theta})=\theta$

设总体X方差σ^2未知,σ^2的据估计量

$S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{x})^2$是有偏的

$E(S_n^2)=\frac{n-1}{n}\neq\sigma^2$,所以$\hat{\sigma}^2=S_n^2$是有偏的。所以修正样本方差$\frac{n}{n-1}S_n^2=S_{n-1}^2$是无偏的。

有效性

$\hat{\theta}_1,\hat{\theta}_2$是$theta$的无偏估计量，方差小的较为有效。这里指无偏估计量的方差。若$D(\hat{\theta}_1)\leq D(\hat{\theta}_2)$,则称$\hat{\theta}_1$较$\hat{\theta}_2$有效。

一致性

$\hat{\theta_n}=\theta(x_1,\hat{...},x_n),\lim_{n\to\infty}\hat{\theta_n}\to\theta$

$\lim_{n\to\infty}P(|\hat{\theta_n}-\theta|<\epsilon)=1$

样本k阶矩是总体k阶矩的一致性估计量（由大数定律证明）

$\frac{1}{n}\sum_{i=1}^{n}X_i^k\to\frac{1}{n}\sum_{i=1}^{n}E(X_i^k)=E(X^k)$

设θ_n是θ的无偏估计量，且$lim_{n\to\infty}D(\hat{\theta_n})=0$,则$\hat{\theta}$是$\theta$的一致估计量

矩法得到的估计量一般为一致估计量

区间估计

区间估计：根据样本给出未知参数的一个范围，并保证真参数以指定的较大概率属于这个范围。$P(\hat{\theta_1} < \theta < \hat{\theta_2}) = 1 - \alpha$

置信区间与置信度

定义：设总体 含未知参数 θ；对于样本$X_1, ..., X_n$找出统计量:

$$\hat{\theta_i} = \theta_i(X_1, .., X_n), (i = 1, 2), \hat{\theta_1} < \hat{\theta_2}$$

使得$P(\hat{\theta_1} < \theta < \hat{\theta_2}) = 1 - \alpha$，$0 < \alpha < 1$

称区间$[\hat{\theta_1}, \hat{\theta_2}]$为$\theta$的 置信区间 ，$1 - \alpha$为该区间的 置信度 。

正态总体，求均值的μ区间估计

已知方差，估计均值

未知方差，估计均值

正态总体，求方差σ^2的区间估计

$$\chi = \frac{nS_n^2}{\sigma^2}\ \chi^2(n-1)$$

使概率对称 $P(\chi^2 < \lambda_1) = P(\chi^2 > \lambda_2) = \frac{\alpha}{2}$

$$\chi^2_{1-\frac{\alpha}{2}} \le \frac{nS_n^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2}}(n)$$

双正态总体情形

使用的是修正的样本方差S_{n-1}^2

求$\mu_1 - \mu_2, \frac{\sigma_1^2}{\sigma_2^2}$的区间估计。

σ_1^2, σ_2^2已知，求μ_1, μ_2的置信区间

$$\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n_1}), \bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$$

$$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

化为标准正态分布后查表

方差比$\frac{\sigma_1^2}{\sigma_2^2}$的置信区间

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

置信区间$(\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1-1, n_2-1)}, \frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n_1-1, n_2-1)})$

单侧置信区间

在单侧置信区间中，都是分位点都是α

对$0 < \alpha < 1$,样本$X_1, .., X_n$,确定统计量$\hat{\theta}(X_1, ..., X_n)$使$P(\theta > \hat{\theta_1}) = 1 - \alpha$,则称$(\hat{\theta_1}, +\infty)$是$\theta$的置信度$1 - \alpha$的单侧置信区间，$\hat{\theta}$称为单侧置信下限。

类似有$P(\theta < \hat{\theta_2}) = 1 - \alpha$，位单侧置信上限。

例如$X \sim N(\mu, \sigma^2)$，求μ的单侧置信下限，$T = \frac{\bar{X} - \mu}{S_{n-1}/\sqrt{n}} \sim t(n-1)$