第五章

第一节 统计量及其分布

- 简单随机样本与统计量
- 经验分布函数
- 样本均值的分布
- 中心极限定理

一、简单随机样本及其分布

1. 样本与样本观测值

样本：按一定的规定从总体X中抽出的一部分个体.

由于抽到哪一些个体作为样本是随机的，因此样本是随机变量，用$X_1,X_2,\dots,X_n$表示；

样本在抽取之后，经观测就有确定的观测值，用$x_1,x_2,\dots,x_n$表示

样本容量：样本中所包括的个体数n

一、简单随机样本及其分布

2. 简单随机样本

设$X_1,X_2,\dots,X_n$为来自总体X的样本，如果满足：

(1) $X_1,X_2,\dots,X_n$相互独立；

(2) $X_1,X_2,\dots,X_n$与X具有相同的分布；

则称$X_1,X_2,\dots,X_n$为总体X的简单随机样本，以下简称样本。

注　大样本；小样本

一、简单随机样本及其分布

3. 简单随机样本的分布

若总体$X\sim F(x)$，则样本$(X_1,X_2,\cdots,X_n)$的分布函数为

$$F^*(x_1,x_2,\dots,x_n)=\prod_{i=1}^{n}F(x_i)$$

确定样本的分布函数的作用？

一、简单随机样本及其分布

1）离散型

若总体X有分布律 $P\{X=x_k\}=p_k, k=1,2,..$

则样本$(X_1,X_2,\dots,X_n)$的联合分布律为

$$P\{X_1=a_1,X_2=a_2,\dots,X_n=a_n\}=\prod_{i=1}^{n}P\{X=a_i\}$$

2）连续型

设$X\sim f(x)$，则样本$(X_1,X_2,\dots,X_n)$的概率密度

$$f(x_1,x_2,\dots,x_n)=\prod_{i=1}^{n}f(x_i)$$

一、简单随机样本及其分布

例 设总体$X\sim(0\text{-}1)$分布，$0<p<1$

$$P\{X=x\}=p^{x}(1-p)^{1-x}, x=0,1$$

$$P\{X_1=x_1,X_2=x_2,\ldots,X_n=x_n\}$$

$$=\prod_{i=1}^{n}P\{X_i=x_i\}=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}$$

$$=p^{\sum\limits_{i=1}^{n}x_i}(1-p)^{n-\sum\limits_{i=1}^{n}x_i},\quad x_i=0,1$$

一、简单随机样本及其分布

例 总体$X\sim Z(\alpha)$, i.e. $Exp(\alpha)$

$$f(x_1,x_2,\ldots,x_n)=\prod_{i=1}^{n}f(x_i)$$

$$=\begin{cases}\prod\limits_{i=1}^{n}\left(\alpha e^{-\alpha x_i}\right)=\alpha^n e^{-\alpha\sum\limits_{i=1}^{n}x_i} & x_i>0, i=1,\ldots,n\\ 0 & \text{其 它}\end{cases}$$

二、统计量

1. 统计量的概念

设$(X_1,X_2,\ldots,X_n)$为总体X的一个样本,

不含任何未知参数的样本的函数

$$T=T(X_1,X_2,\ldots,X_n)$$

称为统计量。

$$t=T(x_1,x_2,\ldots,x_n)$$

称为统计值。

课 堂 练 习

设 $X_1,X_2,\ldots,X_n$ i.i.d. $X\sim N(\mu,\sigma^2)$

其中μ 已知，σ 未知，则

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

—— 统计量

$$\frac{X_1}{\sigma}+\frac{X_2}{\sigma} \qquad \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$$

—— 不是统计量

二、统计量

2. 顺序统计量

$$X_{(1)}\le X_{(2)}\le\ldots\le X_{(n)}$$

$$\begin{cases}X_{(1)}=\min\{X_1,X_2,\ldots,X_n\}\\ X_{(n)}=\max\{X_1,X_2,\ldots,X_n\}\end{cases}$$

若已知总体$X\sim F(x)$，则

(1) $X_{(1)}\sim F_{X_{(1)}}(x)=1-(1-F(x))^n$

(2) $X_{(n)}\sim F_{X_{(n)}}(x)=[F(x)]^n$

二、统计量

3. 常见统计量

(1) 样本均值 $\overline{X}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$

(2) 样本中位数 $M_e=\begin{cases}X_{(k+1)} & n=2k+1\\ \frac{1}{2}\left(X_{(k)}+X_{(k+1)}\right) & n=2k\end{cases}$

(3) 样本极差 $D_n=X_{(n)}-X_{(1)}$

(4) 样本方差 $S^2=\frac{1}{n-1}\sum\limits_{i=1}^{n}\left(X_i-\overline{X}\right)^2$

(5) 样本标准差 $S=\sqrt{S^2}$

二、统计量

4. 样本矩

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

（2） $S_n^k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$

变异系数

三、样本均值的分布 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

已知$X_1,X_2,\cdots,X_n$ i.i.d. X，且$E(X)=\mu, D(X)=\sigma^2$

$E(\overline{X})=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\mu$　　$D(\overline{X})=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)=\frac{\sigma^2}{n}$

1. 若$X\sim N(\mu,\sigma^2)$，则　$\overline{X}\sim N(\mu,\ \sigma^2/n)$

2. 若$X\sim F(x)$，$\overline{X}\sim ?$　$\overline{X}\xrightarrow[n\to\infty]{\text{近似}}N\left(\mu,\frac{\sigma^2}{n}\right)$

四、中心极限定理

1. 基本概念

设r. v. $X_1,X_2,\cdots$的部分和为$Y_n=\sum_{k=1}^{n}X_k$，若有

$$\frac{Y_n-E(Y_n)}{\sqrt{D(Y_n)}}\xrightarrow{n\to\infty}N(0,1)$$

则称$X_1,X_2,\cdots$服从中心极限定理。

在什么条件下，随机变量和的分布函数会收敛于正态分布？—— 中心极限定理（CLT）

四、中心极限定理

2. 独立同分布CLT

若随机变量序列$X_1,X_2,\cdots\cdots$独立同分布，

且 $E(X_k)=\mu,\ D(X_k)=\sigma^2>0$　　$k=1,2,\cdots$

记 $Y_n=\sum_{k=1}^{n}X_k$，则

$$Y_n^*=\frac{Y_n-n\mu}{\sqrt{n}\sigma}\xrightarrow{\text{近似}}N(0,1)$$

四、中心极限定理

例1. 每盒巧克力的净重为随机变量，平均重量为100克，标准差为10克. 一箱内装200盒巧克力，求一箱巧克力的净重大于20500克的概率?

例2. 为了解某电视节目在成都市的收视率p，调查公司随机调查200个成都市民，并将这200个调查对象中收看此节目的频率作为真实收视率的估计。试问调查所得的收视频率与真实收视率p之间的差异不大于5%的可能性有多大？

课堂练习

某药厂生产的某种药品，声称对某疾病的治愈率为80%，现为了检验此治愈率，任意抽取100个此种病患者进行临床试验。如果有多于75人治愈,则此药通过检验。试在以下两种情况下，分别计算此药通过检验的可能性:

(1) 此药的实际治愈率为80%;

(2) 此药的实际治愈率为70%.

解: 记Y_n=100个此种病患者中治愈的人数，则

(1) $Y_n \sim B(100, 0.8) \;\Rightarrow E(Y_n) = 80 \;\; Var(Y_n) = 16$

$$P\{Y_n \geq 75\} \approx 1 - \Phi\left(\frac{75-80}{4}\right) = 1 - \Phi\left(-\frac{5}{4}\right)$$

$$= \Phi(1.25) = ??$$

此药通过检验的可能性是很大的

(2) $Y_n \sim B(100, 0.7)$

$$\Rightarrow E(Y_n) = 70 \;\; Var(Y_n) = 21$$

$$P\{Y_n \geq 75\} \approx 1 - \Phi\left(\frac{75-70}{\sqrt{21}}\right) = 1 - \Phi(1.09) = ??$$

可见，此药通过检验的可能性是很小的

例 题 -3

在总体$X \sim N(52, 6.3^2)$中随机抽一容量为36的样本，试求$P\left(50.8 < \overline{X} < 53.8\right)$。

解：$\left.\begin{array}{r} X \sim N(52, 6.3^2) \\ n = 36 \end{array}\right\} \Rightarrow \overline{X} \sim N\left(52, \frac{6.3^2}{36}\right)$

$$P\{50.8 < \overline{X} < 53.8\} = \Phi\left(\frac{53.8-52}{1.05}\right) - \Phi\left(\frac{50.8-52}{1.05}\right)$$

$$= \Phi(1.71) - \Phi(-1.41) = 0.8293$$

例 题 -4

总体$X \sim N(1, 0.2^2)$,从中抽取容量为n的样本,欲使 $P\{0.9 \leq \overline{X} < 1.1\} \geq 0.95$

试求样本容量n最小应取多大?

解：$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(1, \frac{0.2^2}{n}\right)$

$$P\{0.9 \leq \overline{X} < 1.1\} = 2\Phi\left(\frac{\sqrt{n}}{2}\right) - 1 \geq 0.95$$

$$\Rightarrow n \geq 15.3664$$

例 题 -5

设总体$X \sim U\left(-\frac{1}{2}, 1\right)$, $X_1, X_2, \cdots X_n$ 是来自该总体的简单随机样本。试求:

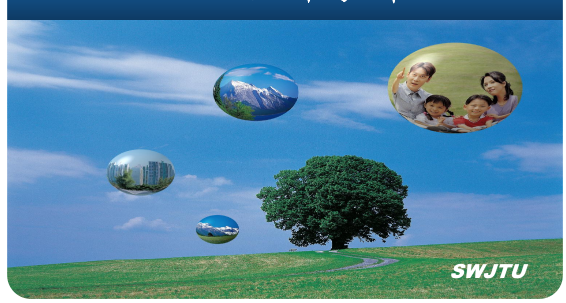
(1) $P\left\{X_{(1)} > 0\right\}$;

(2) $P\left\{X_{(n)} > 0\right\}$.

第四章

第二节

三大抽样分布

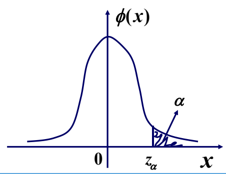


一、上α- 分位点

定义　设随机变量$X\sim F(x)$，$\forall\alpha(0<\alpha<1)$，若

$$1-F(\omega_\alpha)=P\{X\geq\omega_\alpha\}=\alpha$$

则称此实数ω_α为分布$F(x)$的上 $\alpha-$分位点。



例如：$z_{0.05}=1.645$

$z_{0.005}=2.575$

二、三大抽样分布

1.　χ^2 分布

$$X_1,X_2,\cdots,X_n\ i.i.d.\ N(0,1)\ \Rightarrow\chi_n^2=\sum_{i=1}^n X_i^2\sim\chi^2(n)$$

$$f(x)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2} & x>0\\ 0 & x\leq 0\end{cases}$$

注：$\Gamma(x)=\int_0^{+\infty}t^{x-1}e^{-t}dt$

$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)\quad\Gamma(n+1)=n!$

二、三大抽样分布

2.　χ^2 分布的性质

(1)　$E\left[\chi^2(n)\right]=n\qquad D\left[\chi^2(n)\right]=2n$

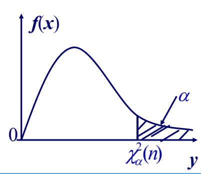
(2)　$\left.\begin{array}{l}X\sim\chi^2(m),\ Y\sim\chi^2(n)\\ X与Y相互独立\end{array}\right\}\Rightarrow\chi_1^2+\chi_2^2\sim\chi^2(m+n)$

$\forall 0<\alpha<1,\ P\left\{\chi^2(n)>\chi_\alpha^2(n)\right\}=\alpha$

——上$\alpha-$分位点

注：$n>45$时，

$$\chi_\alpha^2(n)\approx\frac{1}{2}\left(z_\alpha+\sqrt{2n-1}\right)^2$$



二、三大抽样分布

2.　t 分布

$$\left.\begin{array}{l}X\sim N(0,1),\ Y\sim\chi^2(n)\\ X与Y相互独立\end{array}\right\}\Rightarrow T=\frac{X}{\sqrt{Y/n}}\sim t(n)$$

(1)　$f(-x)=f(x)$

(2)　$\lim\limits_{n\to\infty}f_{t(n)}(x)=\varphi(x)$　　　上$\alpha-$分位点

(3)　$E[t(n)]=0$　　　注1：　$t_{1-\alpha}(n)=-t_\alpha(n)$

$D[t(n)]=\dfrac{n}{n-2}$　　　注2：　$n>45$时，$t_\alpha(n)\approx z_\alpha$

二、三大抽样分布

3.　F 分布

$$\left.\begin{array}{l}X\sim\chi^2(m),Y\sim\chi^2(n)\\ X和Y相互独立\end{array}\right\}\Rightarrow F=\frac{X/m}{Y/n}\sim F(m,n)$$

(1)　$F\sim F(m,n)\Rightarrow\dfrac{1}{F}\sim F(n,m)$

(2)　$E[F(m,n)]=\dfrac{n}{n-2}\quad n>2\qquad D[F(m,n)]=\dfrac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}\quad n>4$

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

三、单个正态总体的抽样分布

设$X_1,X_2,\cdots,X_n\ i.i.d.\ N(\mu,\sigma^2)$，则

1^0　$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$　　　2^0　$\boxed{\dfrac{(n-1)S^2}{\sigma^2}}\sim\chi^2(n-1)$

3^0　$\overline{X}$与S^2相互独立　　　4^0　$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

5^0　$\boxed{\dfrac{1}{\sigma^2}\sum_{i=1}^n(X_i-\mu)^2}\sim\chi^2(n)$　　　$\boxed{\sum_{i=1}^n(X_i-\overline{X})^2=(n-1)S^2}$

四、两个正态总体的抽样分布

$X_1,X_2,\cdots,X_m$ *i.i.d.* $N(\mu_1,\sigma_1^2)$, $Y_1,Y_2,\cdots,Y_n$ *i.i.d.* $N(\mu_2,\sigma_2^2)$

且$X_1,\cdots,X_m,Y_1,\cdots,Y_n$相互独立，则

1^0 $\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}}\sim N(0,1)$

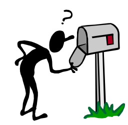
2^0 当$\sigma_1=\sigma_2$时，记 $S_w^2=\dfrac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}$

$$\Rightarrow T=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}\sim t(m+n-2)$$

四、两个正态总体的抽样分布

3^0 $F=\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(m-1,n-1)$

4^0 $\dfrac{n\sigma_2^2\sum\limits_{i=1}^{m}(X_i-\mu_1)^2}{m\sigma_1^2\sum\limits_{i=1}^{n}(Y_i-\mu_2)^2}\sim F(m,n)$



例1. 设$r.v.X\sim t(n)$,试证：$X^2\sim F(1,n)$.

例2. 设(X_1,X_2)是来自正态总体$N(0,\sigma^2)$的一个简单随机样本。

（1）证明：X_1+X_2与X_1-X_2相互独立；

(2) 计算 $P\left\{\dfrac{(X_1+X_2)^2}{(X_1-X_2)^2}<40\right\}$.