概率论与数理统计笔记

刘承杰

南京大学软件学院

2021 年 8 月 15 日

目录

1 概率论的基本概念

1.1 随机试验

定义 1.1 (随机试验) 在概率论中，我们将具有以下三个特点的试验称为随机试验：

1. 可以在相同的条件下重复的进行；
2. 每次试验的可能结果不止一个，并且能事先明确试验的所有可能结果；
3. 进行一次试验之前不能确定哪个结果会出现

1.2 样本空间、随机事件

定义 1.2 (样本空间) 随机事件E的所有可能结果组成的集合，记为S

定义 1.3 (样本点) 样本空间的元素，即E的每个结果

定义 1.4 (随机事件) 试验E的样本空间S的子集为E的随机事件，简称**事件**. 在每次试验中，当且仅当这一子集中的一个样本点出现时，称这一**事件发生**

定义 1.5 (基本事件) 由一个样本点组成的单点集，称为基本事件

定义 1.6 (必然事件和不可能事件) 样本空间S包含所有的样本点，它是S自身的子集，在每次试验中它总是发生的，S称为 **必然事件**.空集$\emptyset$不包含任何样本点，它也作为样本空间的子集，它在每次试验中都不可能发生，$\emptyset$称为**不可能事件**.

定义 1.7 (事件间的关系与事件的运算) 设试验E的样本空间为S,而$A,B,A_k(k=1,2,\cdots)$是S的子集.

1. 若$A\subset B$,则称事件B**包含**事件A,这指的是事件A发生，则事件B必然发生。若$A\subset B$且$B\subset A$,即$A=B$,则称事件A与事件B**相等**。
2. 事件$A\cup B=\{x|x\in A$或$x\in B\}$称为事件A与事件B的**和事件**。当且仅当A,B中至少有一个发生时，事件$A\cup B$发生。
   类似地，称$\bigcup\limits_{k=1}^{n}A_k$为$n$个事件$A_1,A_2,\cdots,A_n$的和事件；称$\bigcup\limits_{k=1}^{\infty}A_k$为可列个事件$A_1,A_2,\cdots$的和事件
3. 事件$A\cap B=\{x|x\in A$且$x\in B\}$称为事件A与事件B的**积事件**。当且仅当A,B同时发生时，事件$A\cap B$发生。
   类似地，称$\bigcap\limits_{k=1}^{n}A_k$为$n$个事件$A_1,A_2,\cdots,A_n$的积事件；称$\bigcap\limits_{k=1}^{\infty}A_k$为可列个事件$A_1,A_2,\cdots$的积事件
4. 事件$A-B=\{x|x\in A$且$x\notin B\}$称为事件A与事件B的**差事件**。当且仅当 A发生、B不发生时事件$A-B$发生。
5. 若$A\cap B=\emptyset$,则称事件A与B是互不相容的，或**互斥的**。这指的是事件A与B 不能同时发生。基本事件是两两互不相容的。

6. 若$A\cup B=S$且$A\cap B=\emptyset$,则称事件A与事件B互为**逆事件**，又称事件A与事件B 互为**对立事件**。这指的是对每次试验而言，事件A、B中必有一个发生，且仅有一个发生。A的对立事件记为$\overline{A}$

定理 1.1 (集合运算定律) 设A、B、C为事件，则有：

交换律：

$$A\cup B=B\cup A$$

$$A\cap B=B\cap A$$

结合律：

$$A\cup(B\cup C)=(A\cup B)\cup C$$

$$A\cap(B\cap C)=(A\cap B)\cap C$$

分配律：

$$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$$

$$A\cap(B\cup C)=(A\cap B)\cap(A\cap C)$$

德摩根律：

$$\overline{A\cup B}=\overline{A}\cap\overline{B}$$

$$\overline{A\cap B}=\overline{A}\cup\overline{B}$$

1.3 频率与概率

定义 1.8 (频率) 在相同的条件下，进行了n次试验，在这n次试验中，事件A发生的次数n_A称为事件A发生的频数。比值$\frac{n_A}{n}$称为事件A发生的频率，并记为$f_n(A)$.

定义 1.9 (概率) 设E是随机试验，S是它的样本空间，对于E的每一个事件A赋予一个实数，记为$P(A)$，称为事件A的概率，如果集合函数$P(\cdot)$满足下列条件：

1. **非负性**：对于每一个事件A，有$P(A)\geq 0$;
2. **规范性**：对于必然事件S,有$P(S)=1$;
3. **可列可加性**：设$A_1,A_2,\cdots$是两两不相容的事件，即对于$A_iA_j=\emptyset,i\neq j,i,j=1,2,\cdots$,有

$$P(A_1\cup A_2\cup\cdots)=P(A_1)+P(A_2)+\cdots$$

当$n\to\infty$时频率$f_n(A)$在一定意义下接近于概率$P(A)$

1.4 等可能概型（古典概型）

定义 1.10 (等可能概型) 具有以下两个特点的试验称为等可能概型：

1. 试验的样本空间只包含有限个元素；
2. 试验中每个基本事件发生的可能性相同。

定理 1.2 (等可能概型中事件A的概率计算公式) 若事件A包含K个基本事件，即$A = e_{i_1} \cup e_{i_2} \cup \cdots \cup e_{i_k}$，这里$i_1, i_2, \cdots, i_k$是$1, 2 \cdots, n$中某$k$个不同的数，则有：

$$P(A) = \sum_{j=1}^{k} P(\{e_{i_j}\}) = \frac{k}{n} = \frac{A\text{包含的基本事件数}}{S\text{中基本事件的总数}}$$

定理 1.3 (超几何分布的概率公式) 设共有N件产品，其中有D件次品，从中取n件，其中恰好有$k(k \leq D)$件次品的概率为：

$$p = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}$$

1.5 条件概率

定义 1.11 (条件概率) 设A, B是两个事件，且$P(A) > 0$，称

$$P(B|A) = \frac{P(AB)}{P(A)}$$

为在事件A发生的条件下事件B发生的条件概率。

定理 1.4 (乘法定理) 设$P(A) > 0$,则有

$$P(AB) = P(B|A)P(A)$$

上式可以推广到多个事件的积事件的情况。例如，设A, B, C为事件，且$P(AB) > 0$,则有

$$P(ABC) = P(C|AB)P(B|A)P(A)$$

一般地，设$A_1, A_2, \cdots, A_n$为n个事件，$n \geq 2$,且$P(A_1A_2\cdots A_{n-1}) > 0$,则有

$$P(A_1A_2\cdots A_n) = P(A_n|A_1A_2\cdots A_{n-1})P(A_{n-1}|A_1A_2\cdots A_{n-2})\cdots P(A_2|A_1)P(A_1)$$

定义 1.12 (样本空间的划分) 设S为试验E的样本空间，$B_1, B_2, \cdots, B_n$为E的一组事件，若

1. $B_iB_j = \emptyset, i \neq j, i, j = 1, 2, \cdots, n$
2. $B_1 \cup B_2 \cup \cdots \cup B_n = S$

则称$B_1, B_2, \cdots, B_n$为样本空间的一个划分。

若$B_1, B_2, \cdots, B_n$为样本空间的一个划分，那么对于每次试验，事件$B_1, B_2, \cdots, B_n$中必有且只有一个发生。

定理 1.5 (全概率公式) 设试验E的样本空间为S，A为E的事件，$B_1, B_2, \cdots, B_n$为S的一个划分，且$P(B_i) > 0, i = 1, 2, \cdots, n$, 则

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)$$

定理 1.6 (贝叶斯(Bayes)公式) 设试验E的样本空间为S,A为E的事件，$B_1, B_2, \cdots, B_n$为S的一个划分，且$P(A) > 0, P(B_i) > 0, i = 1, 2, \cdots, n$,则

$$P(B_i|A) = \frac{P(B_iA)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum\limits_{j=1}^{n} P(A|B_j)P(B_j)}, i = 1, 2, \cdots, n$$

1.6 独立性

定义 1.13 (独立性) 设A, B是两事件，如果满足等式

$$P(AB) = P(A)P(B)$$

则称事件A, B相互独立

同理，对于A, B, C三个事件，如果满足等式

$$P(AB) = P(A)P(B)$$

$$P(BC) = P(B)P(C)$$

$$P(AC) = P(A)P(C)$$

$$P(ABC) = P(A)P(B)P(C)$$

则称事件A, B, C相互独立。

一般地，设$A_1, A_2, \cdots, A_n$是$n(n \geq 2)$个事件，如果对于其中任意2个，任意3个，$\cdots$,任意n个事件的积事件的概率都等于各事件概率的积，则称事件$A_1, A_2, \cdots, A_n$相互独立。

2 随机变量及其分布

2.1 随机变量

定义 2.1 (随机变量) 设随机试验的样本空间为$S=\{e\}$,$X=X(e)$是定义在样本空间S上的实值单值函数，称$X=X(e)$为随机变量

2.2 离散型随机变量及其分布律

定义 2.2 (离散型随机变量) 全部可能取到的值是有限个或可列无限多个的随机变量，称为离散型随机变量。

定义 2.3 ((0 − 1)分布) 设随机变量X只能取$0,1$两个值，它的分布律是

$$P\{X=k\}=p^k(1-p)^{1-k}, k=0,1 \quad (0<p<1)$$

定义 2.4 (伯努利试验) 设试验E只有两个可能结果：A和$\overline{A}$,则称E为伯努利(Bernoulli)试验.设$P(A)=p(0<p<1)$, 此时$P(\overline{A})=1-p$,将E独立重复进行n次，则称这一串重复的独立试验为n重伯努利试验。

定义 2.5 (二项分布) 以X表示n重伯努利试验中事件A发生的次数，事件A在指定的$k(0\le k\le n)$次试验中发生的概率为

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

随机变量X服从参数为n,p的二项分布，并记为$X\sim b(n,p)$.

定义 2.6 (泊松分布) 设随机变量X所有可能取的值为$0,1,2,\cdots$,而取各个值的概率为

$$P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!}, k=0,1,2,\cdots,$$

其中$\lambda>0$是常数。则称X服从参数为λ的泊松分布，记为$X\sim\pi(\lambda)$

定理 2.1 (泊松定理) 设$\lambda>0$是一个常数，n是任意正整数，设$np_n=\lambda$,则对于任一固定的非负整数k,有

$$\lim_{n\to\infty}\binom{n}{k}{p_n}^k(1-p_n)^{n-k}=\frac{\lambda^k e^{-\lambda}}{k!}$$

2.3 随机变量的分布函数

定义 2.7 (随机变量的分布函数) 设X是一个随机变量，x是任意实数，函数

$$F(X)=P\{X\le x\}, -\infty<x<\infty$$

称为X的分布函数

定理 2.2 (分布函数性质)

1. $F(x)$是一个不减函数
2. $0 \leq x \leq 1$,且$F(-\infty) = \lim_{x\to-\infty} F(x) = 0, F(\infty) = \lim_{x\to\infty} F(x) = 1$
3. $F(x+0) = F(x)$,即$F(x)$是右连续的

反之，具备上述三条性质的函数必是某个随机变量的分布函数。

2.4 连续型随机变量及其概率密度

定义 2.8 (连续型随机变量、概率密度) 如果对于随机变量X的分布函数$F(x)$，存在非负可积函数$f(x)$，使对于任意实数x有

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

则称X为**连续型随机变量**，$f(x)$称为X的概率密度函数，简称**概率密度**。

定义 2.9 (均匀分布) 若连续型随机变量X具有概率密度

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

则称X在区间(a,b)上服从均匀分布，记为$X \sim U(a,b)$

X的分布函数为

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$$

定义 2.10 (指数分布) 若连续型随机变量X的概率密度为

$$f(x) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & \text{其他} \end{cases}$$

其中$\theta > 0$为常数，则称X服从参数为θ的指数分布

X的分布函数为

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & \text{其他} \end{cases}$$

定义 2.11 (正态分布) 若连续型随机变量X的概率密度为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

其中$\mu, \sigma(\sigma > 0)$为常数，则称X服从参数为μ, σ的正态分布或高斯分布，记为$X \sim N(\mu, \sigma^2)$

X的分布函数为

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt$$

特别地，当$\mu=0,\sigma=1$时，称随机变量X服从**标准正态分布**，其概率密度和分布函数分别用$\varphi(x),\Phi(x)$表示，有

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}\,dt$$

对于一般的正态分布，只需通过一个线性变换就能化为标准正态分布：

$$\text{若}X\sim N(\mu,\sigma^2),\text{则}Z=\frac{X-\mu}{\sigma}\sim N(0,1)$$

2.5 随机变量的函数的分布

定理 2.3 设随机变量X具有概率密度$f_x(x),-\infty<x<\infty$,又设函数$g(x)$处处可导且恒有$g'(x)>0$（或恒有$g'(x)<0$，则$Y=g(X)$是连续型随机变量，其概率密度为

$$f_Y(y)=\begin{cases} f_x[h(y)]|h'(y)|, & \alpha<y<\beta \\ 0, & \text{其他} \end{cases}$$

其中$\alpha=\min\{g(-\infty),g(\infty)\},\beta=\max\{g(-\infty),g(\infty)\}$,$h(y)$是$g(x)$的反函数

3 多维随机变量及其分布

3.1 二维随机变量

定义 3.1 (联合分布函数) 设(X,Y)是二维随机变量，对于任意实数x,y，二元函数

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

称为二维随机变量(X,Y)的**分布函数**，或称为随机变量X和Y的**联合分布函数**。

如果将二维随机变量(X,Y)看成是平面上随机点的坐标，那么分布函数$F(x,y)$在(x,y)处的函数值就是随机点 (X,Y)落在以点(x,y)为顶点而位于该点左下方的无穷矩形区域内的概率。所以，随机点(X,Y)落在矩形区域$\{(x,y)|x_1 < x \le x_2, y_1 < y \le y_2\}$的概率为

$$P\{x_1 < x \le x_2, y_1 < y \le y_2\} = F(x_2,y_2) - F(x_2,y_1) + F(x_1,y_1) - F(x_1,y_2)$$

定义 3.2 (联合概率密度) 对于二维随机变量(X,Y)的分布函数$F(x,y)$，如果存在非负可积函数$f(x,y)$使对于任意x,y有

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

则称(X,Y)是**连续型的二维随机变量**,函数$f(x,y)$称为二维随机变量(X,Y)的**概率密度**，或称为随机变量X和Y的**联合概率密度**。

3.2 边缘分布

定义 3.3 (边缘分布函数) 二维随机变量(X,Y)作为一个整体，具有分布函数$F(x,y)$，而X和Y都是随机变量，各自也有分布函数，将他们分别记为 $F_X(x), F_Y(y)$,依次称为二维随机变量(X,Y)关于X和关于Y的**边缘分布函数**,且

$$F_X(x) = F(x,\infty), F_Y(y) = F(\infty,y)$$

定义 3.4 (边缘分布律) 记

$$p_{i\cdot} = \sum_{j=1}^{\infty} p_{ij} = P\{X = x_i\}, \quad i = 1,2,\cdots,$$

$$p_{\cdot j} = \sum_{i=1}^{\infty} p_{ij} = P\{Y = y_j\}, \quad j = 1,2,\cdots,$$

分别称$p_{i\cdot}(i = 1,2,\cdots)$和$p_{\cdot j}(j = 1,2,\cdots)$为$(X,Y)$关于$X$和$Y$的**边缘分布律**

定义 3.5 (边缘概率密度) 记

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

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

分别称$f_X(x), f_Y(y)$为(X,Y)关于X和Y的边缘概率密度。

3.3 条件分布

定义 3.6 (条件分布律) 设(X,Y)是二维离散型随机变量，对于固定的j，若$P\{Y=y_j\}>0$,则称

$$P\{X=x_i|Y=y_j\}=\frac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=\frac{p_{ij}}{p_{\cdot j}},i=1,2,\cdots$$

为在$Y=y_j$条件下随机变量X的条件分布律

同样，对于固定的i，若$P\{X=x_i\}>0$,则称

$$P\{Y=y_j|X=x_i\}=\frac{P\{X=x_i,Y=y_j\}}{P\{X=X_i\}}=\frac{p_{ij}}{p_{i\cdot}},j=1,2,\cdots$$

为在$X=x_i$条件下随机变量Y的条件分布律

定义 3.7 (条件概率密度) 设二维随机变量(X,Y)的概率密度为$f(x,y)$,(X,Y)关于Y的边缘概率密度为$f_Y(y)$.若对于固定的y, $f_Y(y)>0$，则称$\frac{f(x,y)}{f_Y(y)}$为在$Y=y$的条件下X的条件概率密度，记为

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

称$\int_{-\infty}^{x}f_{X|Y}(x|y)\,dx=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}\,dx$为在$Y=y$的条件下$X$的条件分布函数，记为 $P\{X\leq x|Y=y\}$或$F_{X|Y}(x|y)$，即

$$F_{X|Y}(x|y)=P\{X\leq x|Y=y\}=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}\,dx$$

类似地，可以定义$f_{Y|X}(y|x)=\dfrac{f(x,y)}{f_X(x)}$和$F_{Y|X}(y|x)=\displaystyle\int_{-\infty}^{y}\frac{f(x,y)}{f_X(x)}\,dy$

3.4 相互独立的随机变量

定义 3.8 (相互独立) 若对于所有x,y，满足下列条件之一，则称随机变量X和Y是相互独立的：

1. 设$F(x,y)$及$F_X(x),F_Y(y)$分别是二维随机变量(X,Y)的分布函数及边缘分布函数，有

$$F(x,y)=F_X(x)F_Y(y)$$

2. 设$f(x,y),f_X(X),f_Y(y)$分别为(X,Y)的概率密度和边缘概率密度,有

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

3. 当X,Y是离散型随机变量是，对于(X,Y)的所有可能取值(x_i,y_i)，有

$$P\{X=x_i,Y=y_i\}=P\{X=x_i\}P\{Y=y_i\}$$

定理 3.1 设$(X_1,X_2,\cdots,X_m)$和$(Y_1,Y_2,\cdots,Y_n)$相互独立，则$X_i(i=1,2,\cdots,m)$和$Y_j(j=1,2,\cdots,n)$相互独立。又若h,g是连续函数，则$h(X_1,X_2,\cdots,X_m)$和$g(Y_1,Y_2,\cdots,Y_n)$相互独立。

3.5 两个随机变量的函数的分布

定义 3.9 ($Z = X + Y$分布) 设(X, Y)是二维连续型随机变量，它具有概率密度$f(x, y)$，则$Z = X + Y$仍为连续型随机变量，其概率密度为

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f(z-y, y)\, dy, \tag{1}$$

或

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f(x, z-x)\, dx. \tag{2}$$

又若X和Y相互独立，设(X, Y)关于X, Y的边缘密度分别为$f_X(x), f_Y(y)$，则(1),(2)分别可化为

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy$$

和

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$$

这两个公式称为f_X和f_Y的**卷积公式**，记为$f_X * f_Y$，即

$$f_X * f_Y = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$$

定理 3.2 有限个相互独立的正态随机变量的线性组合仍然服从正态分布

定义 3.10 ($Z = \dfrac{Y}{X}$分布、$Z = XY$的分布) 设(X, Y)是二维连续型随机变量，它具有概率密度$f(x, y)$，则$Z = \dfrac{Y}{X}, Z = XY$仍为连续型随机变量，其概率密度分别为

$$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f(x, xz)\, dx; \tag{3}$$

$$f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f(x, \frac{z}{x})\, dx. \tag{4}$$

又若X和Y相互独立，设(X, Y)关于X, Y的边缘密度分别为$f_X(x), f_Y(y)$，则(3)可化为

$$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz)\, dx$$

(4)可化为

$$f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(\frac{z}{x})\, dx$$

定义 3.11 ($M = \max\{X, Y\}$分布及$N = \min\{X, Y\}$分布) 设X, Y是两个相互独立的随机变量，它们的分布函数分别为$F_X(x)$和$F_Y(y)$，可得$M = \max\{X, Y\}$的分布函数为

$$F_{\max}(z) = F_X(z) F_Y(z)$$

类似地，可得到$N=\min\{X,Y\}$的分布函数为

$$F_{\min}(z)=1-[1-F_X(z)][1-F_Y(z)]$$

以上结果容易推广到n个相互独立的随机变量的情况。设$X_1,X_2,\cdots,X_n$是n个相互独立的随机变量，它们的分布函数分别为 $F_{X_i}(x_i)(i=1,2,\cdots,n)$，则$M=\max\{X_1,X_2,\cdots,X_n\}$及$N=\min\{X_1,X_2,\cdots,X_n\}$的分布函数分别为

$$F_{\max}(z)=F_{X_1}(z)F_{X_2}(z)\cdots F_{X_n}(z)$$

$$F_{\min}(z)=1-[1-F_{X_1}(z)][1-F_{X_2}(z)]\cdots[1-FX_n(z)]$$

特别地，当$X_1,X_2,\cdots,X_n$相互独立且具有相同分布函数$F(x)$时有

$$F_{\max}(z)=[F(z)]^n$$

$$F_{\min}(z)=1-[1-F(z)]^n$$

4 随机变量的数字特征

4.1 数学期望

定义 4.1 (数学期望) 设离散型随机变量X的分布律为$P\{X=x_k\},k=1,2,\cdots$.若级数$\sum\limits_{k=1}^{\infty}x_kp_k$绝对收敛，则称级数$\sum\limits_{k=1}^{\infty}x_kp_k$的和为随机变量$X$的**数学期望**，记为$E(X)$，即

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

设连续型随机变量X的概率密度为$f(x)$，若积分$\int_{-\infty}^{\infty}xf(x)\,dx$绝对收敛，则称积分$\int_{-\infty}^{\infty}xf(x)\,dx$的值为随机变量$X$的**数学期望**，记为$E(X)$，即

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

数学期望$E(X)$完全由随机变量X的概率分布所确定。

定理 4.1 设Y是随机变量X的函数：$Y=g(X)$（g是连续函数）

(i) 如果X是离散型随机变量，它的分布律为$P=\{X=x_k\}=p_k,k=1,2,\cdots$,若$\sum\limits_{k=1}^{\infty}g(x_k)p_k$绝对收敛，则有

$$E(Y)=E[g(x)]=\sum_{k=1}^{\infty}g(x_k)p_k$$

(ii) 如果X是连续型随机变量，它的概率密度为$f(x)$，若$\int_{-\infty}^{\infty}g(x)f(x)\,dx$绝对收敛，则有

$$E(Y)=E[g(x)]=\int_{-\infty}^{\infty}g(x)f(x)\,dx$$

上述定理还可以推广到两个或两个以上随机变量的函数的情况。

例如设Z是随机变量X,Y的函数$Z=g(X,Y)$（g是连续函数）,那么，Z是一个一维随机变量。若二维随机变量(X,Y)的概率密度为$f(x,y)$，则有

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

这里设上式右边的积分绝对收敛。又若(X,Y)为离散型随机变量，其分布律为$P=\{X=x_i,Y=y_j\}=p_{ij},i,j=1,2,\cdots$,则有

$$E(Z)=E[g(X,Y)]=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}g(x_i,y_j)p_{ij}$$

这里设上式右边的级数绝对收敛。

定理 4.2 数学期望的性质：

1° 设C是常数，则有$E(C)=C$

2° 设X是一个随机变量，C是常数，则有

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

3° 设X,Y是两个随机变量，则有

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

这一性质可以推广到任意有限个随机变量之和的情况

4° 设X,Y是两个相互独立的随机变量，则有

$$E(XY)=E(X)E(Y)$$

这一性质可以推广到任意有限个相互独立的随机变量之积的情况

4.2 方差

定义 4.2 (方差) 设X是一个随机变量，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$\mathrm{Var}(X)$，即

$$D(X)=\mathrm{Var}(X)=E\{[X-E(X)]^2\}$$

在应用上还引入量$\sqrt{D(X)}$，记为$\sigma(X)$，称为**标准差**或**均方差**。

对于离散型随机变量，有

$$D(X)=\sum_{k=1}^{\infty}[x_k-E(X)]^2p_k$$

其中$P=\{X=x_k\}=p_k,k=1,2,\cdots$是$X$的分布律

对于连续型随机变量，有

$$D(X)=\int_{-\infty}^{\infty}[x-E(X)]^2f(x)\,dx$$

其中$f(x)$是X的概率密度

随机变量X的方差可按下列公式计算

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

定理 4.3 方差的性质：

1° 设C是常数，则$D(C)=0$

2° 设X是随机变量，C是常数，则有

$$D(CX) = C^2 D(X) \qquad D(X+C) = D(X)$$

3° 设X, Y是两个随机变量，则有

$$D(X+Y) = D(X) + D(Y) + 2E\{[X - E(X)][Y - E(Y)]\}$$

特别地，若X, Y相互独立，则有

$$D(X+Y) = D(X) + D(Y)$$

这一性质可以推广到任意有限多个相互独立的随机变量之和的情况

4° $D(X) = 0$的充要条件是X以概率1取常数$E(X)$，即

$$P\{X = E(X)\} = 1$$

定理 4.4 (切比雪夫(Chebyshev)不等式) 设随机变量X具有数学期望$E(X) = \mu$，方差$D(X) = \sigma^2$，则对于任意整数ε，有不等式

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

也可写为

$$P\{|X - \mu| < \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

4.3 协方差及相关系数

定义 4.3 (协方差、相关系数) 量$E\{[X - E(X)][Y - E(Y)]\}$称为随机变量$X$与$Y$的**协方差**，记为$\mathrm{Cov}(X, Y)$，即

$$\mathrm{Cov}(X, Y) = E\{[X - E(X)][Y - E(Y)]\}$$

而

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

称为随机变量的X和Y的**相关系数**。

协方差的计算公式：

$$D(X+Y) = D(X) + D(Y) + 2\mathrm{Cov}(X, Y) \qquad \mathrm{Cov} = E(X, Y) - E(X)E(Y)$$

定理 4.5 协方差的性质：

1. $\mathrm{Cov}(aX, bY) = ab\mathrm{Cov}(X, Y)$，$a, b$是常数；
2. $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$

定理 4.6 ρ_{XY}的性质：

1. $|\rho_{XY}| \le 1$
2. $|\rho_{XY}| = 1$的充要条件是，存在常数a, b，使得$P\{Y = a + bX\} = 1$

ρ_{XY}是一个可以用来表征X, Y之间线性关系紧密程度的量，当$|\rho_{XY}|$较大时，X, Y线性相关程度较好，当$|\rho_{XY}|$较小时，X, Y线性相关程度较差，当$|\rho_{XY}| = 0$时，称X和Y不相关。

4.4 矩、协方差矩阵

定义 4.4 设X和Y是随机变量。
若

$$E(X^k) \quad k = 1, 2, \cdots$$

存在，则称它为X的k**阶原点矩**，简称k**阶矩**.
若

$$E\{[X - E(X)]^k\} \quad k = 2, 3, \cdots$$

存在，则称它为X的k**阶中心矩**.
若

$$E\{[X - E(X)]^k [Y - E(Y)]^l\} \quad k, l = 1, 2, \cdots$$

存在，则称它为X和Y的$k+l$**阶混合中心矩**.

定义 4.5 (协方差矩阵) 先以二维随机变量为例。二维随机变量(X_1, X_2)有四个二阶中心矩（设它们都存在），分别记为

$$c_{11} = E\{[X_1 - E(X_1)]^2\}$$

$$c_{12} = E\{[X_1 - E(X_1)][X_2 - E(X_2)]\}$$

$$c_{21} = E\{[X_2 - E(X_2)][X_1 - E(X_1)]\}$$

$$c_{22} = E\{[X_2 - E(X_2)]^2\}$$

将它们排成矩阵的形式

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

这个矩阵称为随机变量(X_1, X_2)的**协方差矩阵**。

设n维随机变量$(X_1, X_2, \cdots, X_n)$的二阶混合中心矩$c_{ij} = \mathrm{Cov}(X_i, X_j) = E\{[X_i - E(X_i)][X_j - E(X_j)]\}, i, j = 1, 2, \cdots, n$ 都存在，则称矩阵

$$\boldsymbol{C} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

为n维随机变量$(X_1, X_2, \cdots, X_n)$的**协方差矩阵**，该矩阵是一个对称矩阵。

5 大数定律及中心极限定理

5.1 大数定律

定理 5.1 (弱大数定理(辛钦大数定理)) 设$X_1, X_2, \cdots$是相互独立的，服从同一分布的随机变量序列，且具有数学期望$E(X_k) = \mu(k = 1, 2, \cdots)$。作前$n$个变量的算数平均$\frac{1}{n}\sum_{k=1}^{n} X_k$，则对于任意$\varepsilon > 0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right\} = 1$$

定义 5.1 设$Y_1, Y_2, \cdots, Y_n, \cdots$是一个随机变量序列，$a$是一个常数。若对于任意正数$\varepsilon$，有

$$\lim_{n\to\infty} P\{|Y_n - a| < \varepsilon\} = 1$$

则称序列$Y_1, Y_2, \cdots, Y_n, \cdots$**依概率收敛于**$a$，记为

$$Y_n \xrightarrow{P} a$$

定理 5.2 (弱大数定理(辛钦大数定理)) 设随机变量$X_1, X_2, \cdots$相互独立，服从同一分布，且具有数学期望$E(X_k) = \mu(k = 1, 2, \cdots)$，则序列$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$依概率收敛于$\mu$，即$\overline{X} \xrightarrow{P} a$

定理 5.3 (伯努利大数定理) 设f_A是n次独立重复试验中事件A发生的次数，p是事件A在每次试验中发生的概率，则对于任意正数$\varepsilon > 0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{f_A}{n} - p\right| < \varepsilon\right\} = 1$$

或

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

5.2 中心极限定理

定理 5.4 (独立同分布的中心极限定理) 设随机变量$X_1, X_2, \cdots, X_n, \cdots$相互独立，服从同一分布，且具有数学期望和方差：$E(X_k) = \mu, D(X_k) = \sigma^2 > 0(k = 1, 2, \cdots)$,则随机变量之和$\sum_{k=1}^{n} X_k$的标准化变量

$$Y_n = \frac{\sum_{k=1}^{n} X_k - E(\sum_{k=1}^{n} X_k)}{\sqrt{D(\sum_{k=1}^{n} X_k)}} = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$

的分布函数$F_n(x)$对于任意x满足

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P\left\{\frac{\sum\limits_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}\le x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt=\varPhi(x)$$

评论 该定理表明，均值为μ，方差为$\sigma^2>0$的独立同分布的随机变量$X_1,X_2,\cdots,X_n$之和$\sum\limits_{k=1}^{n}X_k$的标准化变量，当n充分大时，有

$$\frac{\sum\limits_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}\sim N(0,1)$$

也可改写为

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)\quad 或\quad \overline{X}\sim N(\mu,\sigma^2/n)$$

定理 5.5 (李雅普诺夫(Lyapunov)定理) 设随机变量$X_1,X_2,\cdots,X_n,\cdots$相互独立，它们具有数学期望和方差$E(X_k)=\mu,D(X_k)=\sigma_k^2>0,k=1,2,\cdots$，记$B_n^2=\sum\limits_{k=1}^{n}\sigma_k^2$。若存在整数$\delta$，使得当$n\to\infty$时，

$$\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}E\{|X_k-\mu_k|^{2+\delta}\}\to 0,$$

则随机变量之和$\sum\limits_{k=1}^{n}X_k$的标准化变量

$$Z_n=\frac{\sum\limits_{k=1}^{n}X_k-E(\sum\limits_{k=1}^{n}X_k)}{\sqrt{D(\sum\limits_{k=1}^{n}X_k)}}=\frac{\sum\limits_{k=1}^{n}X_k-\sum\limits_{k=1}^{n}\mu_k}{B_n}$$

的分布函数$F_n(x)$对任意x，满足

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P\left\{\frac{\sum\limits_{k=1}^{n}X_k-\sum\limits_{k=1}^{n}\mu_k}{B_n}\le x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt=\varPhi(x)$$

评论 该定理表明，在定理的条件下，随机变量

$$Z_n=\frac{\sum\limits_{k=1}^{n}X_k-\sum\limits_{k=1}^{n}\mu_k}{B_n}$$

当n很大时，近似地服从正态分布$N(0,1)$。由此，当n很大时，$\sum\limits_{k=1}^{n}X_k=B_nZ_n+\sum\limits_{k=1}^{n}\mu_k$ 近似地服从正态分布$N(\sum\limits_{k=1}^{n}\mu_k,B_n^2)$。这也就是说，无论各个随机变量$X_k(k=1,2,\cdots)$服从什么分布，只要满足定理的条件，那么它们的和$\sum\limits_{k=1}^{n}X_k$当$n$很大时，就近似地服从正态分布。

定理 5.6 (棣莫弗-拉普拉斯(De Moivre-Laplace)定理) 设随机变量$\eta_n(n=1,2,\cdots)$服从参数为$n,p(0<p<1)$的二项分布，则对于任意x，有

$$\lim_{n\to\infty} P\left\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \le x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = \Phi(x)$$

评论 该定理表明，正态分布时二项分布的极限分布。当n充分大时，可以用该定理计算二项分布的概率。

6 样本及抽样分布

6.1 随机样本

定义 6.1 设X是具有分布函数F的随机变量，若$X_1, X_2, \cdots, X_n$是具有同一分布函数F的、相互独立的随机变量，则称$X_1, X_2, \cdots, X_n$为从分布函数F（或总体F、或总体X）得到的**容量为n的简单随机样本**，简称**样本**，它们的观察值$x_1, x_2, \cdots, x_n$称为 **样本值**，又称为X的n**个独立的观察值**。

由定义得：若$X_1, X_2, \cdots, X_n$为F的一个样本，则$X_1, X_2, \cdots, X_n$相互独立，且它们的分布函数都是F，所以$(X_1, X_2, \cdots, X_n)$ 的分布函数为

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

又若X具有概率密度f，则$(X_1, X_2, \cdots, X_n)$的概率密度为

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

6.2 直方图和箱线图

直方图和箱线图的定义及画法略。

6.3 抽样分布

定义 6.2

定义 6.3 设$X_1, X_2, \cdots, X_n$是来自总体X的一个样本，$g(X_1, X_2, \cdots, X_n)$是$X_1, X_2, \cdots, X_n$的函数，若g中不含未知参数，则称$g(X_1, X_2, \cdots, X_n)$ 是**一统计量**。

因为$X_1, X_2, \cdots, X_n$都是随机变量，而统计量$g(X_1, X_2, \cdots, X_n)$是随机变量的函数，因此统计量是一个随机变量.设$x_1, x_2, \cdots, x_n$是相应于样本$X_1, X_2, \cdots, X_n$的样本值，则称$g(x_1, x_2, \cdots, x_n)$是$g(X_1, X_2, \cdots, X_n)$的观察值。

下面列出几个常用的统计量，设$X_1, X_2, \cdots, X_n$是来自总体X的一个样本，$x_1, x_2, \cdots, x_n$是这一样本的观察值，定义

样本平均值

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

样本方差

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

样本标准差

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

样本k阶（原点）矩

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,\quad k=1,2,\cdots$$

样本k阶中心矩

$$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k,\quad k=2,3,\cdots$$

它们的观察值分别为

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2=\frac{1}{n-1}(\sum_{i=1}^{n}x_i^2-n\overline{x}^2)$$

$$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

$$a_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k,\quad k=1,2,\cdots$$

$$b_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^k,\quad k=2,3,\cdots$$

定义 6.4 (经验分布函数) 与总体分布函数$F(x)$相应的统计量——经验分布函数的做法为：设X_1, $X_2,\cdots,X_n$是总体F的一个样本，用$S(x),-\infty<x<\infty$表示$X_1,X_2,\cdots,X_n$ 中不大于x的随机变量的个数.定义经验分布函数$F_n(x)$为

$$F_n(x)=\frac{1}{n}S(x),-\infty<x<\infty$$

定义 6.5 (χ^2分布) 设$X_1,X_2,\cdots,X_n$是来自总体$N(0,1)$的样本，则称统计量

$$\chi^2=X_1^2+X_2^2+\cdots+X_n^2$$

服从自由度为n的χ^2分布，记为$\chi^2\sim\chi^2(n)$.此处自由度是指上式右端包含的独立变量的个数。

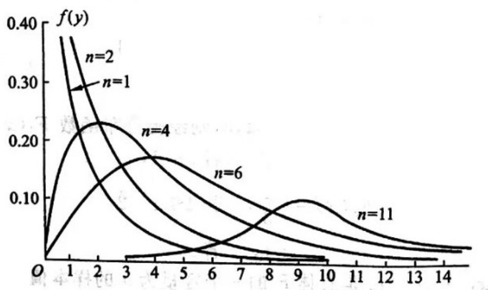

图 1: $\chi^2(n)$分布的概率密度图

$\chi^2(n)$分布的概率密度为

$$f(y)=\begin{cases}\frac{1}{2^{n/2}\Gamma(n/2)}y^{n/2-1}e^{-y/2}, & y>0\\ 0, & \text{其他}\end{cases}$$

$f(y)$的图形如图1所示。

定理 6.1 χ^2分布的性质：

χ^2分布的可加性 设$\chi_1^2\sim\chi^2(n_1),\chi_2^2\sim\chi^2(n_2)$，并且 χ_1^2,χ_2^2相互独立，则有

$$\chi_1^2+\chi_2^2\sim\chi^2(n_1+n_2)$$

χ^2分布的数学期望和方差 若$\chi^2\sim\chi^2(n)$，则有

$$E(\chi^2)=n,\quad D(\chi^2)=2n$$

χ^2分布的上分位点 对于给定的正数$\alpha,0<\alpha<1$，满足条件

$$P\{\chi^2>\chi_\alpha^2(n)\}=\int_{\chi_\alpha^2(n)}^{\infty}f(y)\,dy=\alpha$$

的点$\chi_\alpha^2(n)$就是$\chi^2(n)$分布的上α分位点。

定义 6.6 (t分布) 设$X\sim N(0,1),Y\sim\chi^2(n)$，且$X,Y$相互独立，则称随机变量

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

服从自由度为n的**t分布**.记为$t\sim t(n)$.

t分布又称**学生氏(Student)分布**.$t(n)$分布的概率密度函数为

$$h(t)=\frac{\Gamma[(n+1)/2]}{\sqrt{\pi n}\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2},-\infty<t<\infty$$

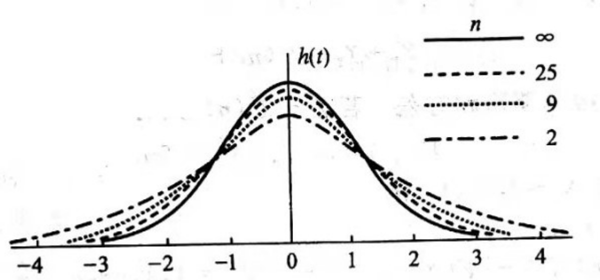

图 2: t分布的概率密度图

定义 6.7 (t分布的上分位点) 对于给定的$\alpha,0<\alpha<1$，满足条件

$$P\{t>t_\alpha(n)\}=\int_{t_\alpha(n)}^{\infty}h(t)\,dt=\alpha$$

的点$t_\alpha(n)$就是$t(n)$分布的上α分位点。由定义及$h(t)$图形的对称性可知

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

定义 6.8 (F分布) 设$U \sim \chi^2(n_1), V \sim chi^2(n_2)$，且$U, V$相互独立，则称随机变量

$$F = \frac{U/n_1}{V/n_2}$$

服从自由度为(n_1, n_2)的F**分布**，记为$F \sim F(n_1, n_2)$.

$F(n_1, n_2)$分布的概率密度为

$$\psi(y) = \begin{cases} \frac{\Gamma[(n_1+n_2)/2](n_1/n_2)^{n_1/2} y^{(n_1/2)-1}}{\Gamma(n_1/2)\Gamma(n_2/2)[1+(n_1 y/n_2)]^{(n_1+n_2)/2}}, & y > 0, \\ 0, & \text{其他} \end{cases}$$

$\psi(y)$的图形如下图所示

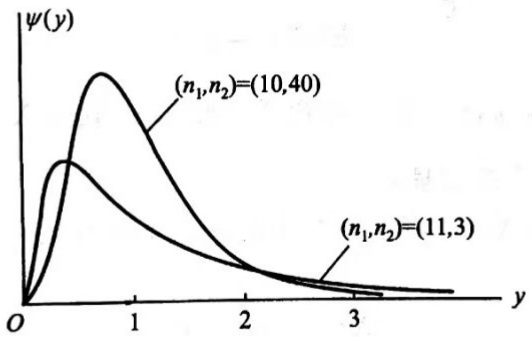

图 3: F分布的概率密度图

由定义可知，若$F \sim F(n_1, n_2)$，则

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

定义 6.9 (F分布的上分位点) 对于给定的$\alpha, 0 < \alpha < 1$，满足条件

$$P\{F > F_\alpha(n_1, n_2)\} = \int_{F_\alpha(n_1,n_2)}^{\infty} \psi(y)\, dy = \alpha$$

的点$F_\alpha(n_1, n_2)$就是$F(n_1, n_2)$分布的上α分位点。

F分布的上α分布有如下的重要性质：

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

定理 6.2 设$X_1, X_2, \cdot, X_n$是来自正态总体$N(\mu, \sigma^2)$的样本，$\overline{X}$是样本均值，则有

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

定理 6.3 设$X_1, X_2, \cdot, X_n$是来自总体$N(\mu, \sigma^2)$的样本，$\overline{X}, S^2$分别是样本均值和样本方差，则有

1° $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$；

2° $\overline{X}$与S^2相互独立

定理 6.4 设$X_1, X_2, \cdot, X_n$是来自总体$N(\mu, \sigma^2)$的样本，$\overline{X}, S^2$分别是样本均值和样本方差，则有

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

定理 6.5 设$X_1, X_2, \cdot, X_{n_1}$与$Y_1, Y_2, \cdot, Y_{n_2}$分别是来自正态总体$N(\mu_1, \sigma_1^2)$和 $N(\mu, \sigma_2^2)$的样本，且这两个样本相互独立.设$\overline{X} = \dfrac{1}{n_1}\sum\limits_{i=1}^{n_1} X_i, \overline{Y} = \dfrac{1}{n_2}\sum\limits_{i=1}^{n_2} Y_i$分别是这两个样本的样本均值；
$S_1^2 = \dfrac{1}{n_1-1}\sum\limits_{i=1}^{n_1}(X_i-\overline{X})^2, S_2^2 = \dfrac{1}{n_2-1}\sum\limits_{i=1}^{n_2}(Y_i-\overline{Y})^2$分别是这两个样本的样本方差，则有

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

2° 当$\sigma_1^2 = \sigma_2^2 = \sigma^2$时，

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2),$$

其中

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

7 参数估计

7.1 点估计

定义 7.1 点估计问题的一般提法如下：设总体X的分布函数$F(x;\theta)$的形式为已知，θ是待估参数.$X_1,X_2,\cdots,X_n$是X的一个样本，$x_1,x_2,\cdots,x_n$是相应的一个样本值. 点估计问题就是要构造一个适当的统计量$\hat{\theta}(X_1,X_2,\cdots,X_n)$，用它的观察值$\hat{\theta}(x_1,x_2,\cdots,x_n)$作为未知参数$\theta$的近似值。称$\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为θ的**估计量**，称$\hat{\theta}(x_1,x_2,\cdots,x_n)$为$\theta$的**估计值**。在不致混淆的情况下统称估计量和估计值为**估计**.

定义 7.2 (矩估计法) 设X为连续型随机变量，其概率密度为$f(x;\theta_1,\theta_2,\cdots,\theta_k)$，或$X$为离散型随机变量，其分布律为$P\{X=x\}=p(x;\theta_1,\theta_2,\cdots,\theta_k)$，其中$\theta_1,\theta_2,\cdots,\theta_k$ 为待估参数，$X_1,X_2,\cdots,X_n$是来自X的样本.假设总体X的前k阶矩

$$\mu_l=E(X^l)=\int_{-\infty}^{\infty}x^l f(x;\theta_1,\theta_2,\cdots,\theta_k)\,dx \quad X\text{为连续型}$$

或

$$\mu_l=E(X^l)=\sum_{x\in R_X}x^l p(x;\theta_1,\theta_2,\cdots,\theta_k) \quad X\text{为离散型}$$

（$l=1,2,\cdots,k$,其中R_X是X可能取值的范围）存在。一般来说，它们是$\theta_1,\theta_2,\cdots,\theta_k$ 的函数.基于样本矩

$$A_l=\frac{1}{n}\sum_{i=1}^{n}X_i^l$$

依概率收敛于相应的总体矩$\mu_l(l=1,2,\cdots,k)$，样本矩的连续函数依概率收敛于相应的总体矩的连续函数，我们就用样本矩作为相应的总体矩的估计量，而以样本矩的连续函数作为相应的总体矩的连续函数的估计量。这种估计方法称为**矩估计法**。

矩估计法的具体做法如下：设

$$\begin{cases}\mu_1=\mu_1(\theta_1,\theta_2,\cdots,\theta_k),\\ \mu_2=\mu_2(\theta_1,\theta_2,\cdots,\theta_k),\\ \vdots\\ \mu_k=\mu_k(\theta_1,\theta_2,\cdots,\theta_k),\end{cases}$$

这是一个包含k个未知参数$\theta_1,\theta_2,\cdots,\theta_k$的联立方程组。一般来说，可以从中解出$\theta_1,\theta_2,\cdots,\theta_k$，得到

$$\begin{cases}\theta_1=\theta_1(\mu_1,\mu_2,\cdots,\mu_k),\\ \theta_2=\theta_2(\mu_1,\mu_2,\cdots,\mu_k),\\ \vdots\\ \theta_k=\theta_k(\mu_1,\mu_2,\cdots,\mu_k),\end{cases}$$

以A_i分别代替上式中的$\mu_i,i=1,2,\cdots,k$，就以

$$\hat{\theta}_i=\theta_i(A_1,A_2,\cdots,A_k),i=1,2,\cdots,k$$

分别作为$\theta_i, i=1,2,\cdots,k$的估计量，这种估计量称为**矩估计量**，矩估计量的观察值称为**矩估计值**.

定义 7.3 (最大似然估计法) 若总体X属离散型，其分布律$P\{X=x\}=p(x;\theta),\theta\in\Theta$的形式为已知，$\theta$为待估参数，$\Theta$是$\theta$可能取值的范围.设$X_1,X_2,\cdots,X_n$是来自$X$的样本，则$X_1,X_2,\cdots,X_n$的联合分布律为

$$\prod_{i=1}^{n} p(x_i;\theta)$$

又设$x_1,x_2,\cdots,x_n$是相应于样本$X_1,X_2,\cdots,X_n$的一个样本值.易知样本$X_1,X_2,\cdots,X_n$取到观察值$x_1,x_2,\cdots,x_n$的概率，亦即事件$\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}$发生的概率为

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

这一概率随θ的取值而变化，它是θ的函数，$L(\theta)$称为样本的**似然函数**。（注意，这里$x_1,x_2,\cdots,x_n$是已知的样本值，它们都是常数）.

由费希尔引进的最大似然估计法，就是固定样本观察值$x_1,x_2,\cdots,x_n$，在θ取值的可能范围Θ内挑选使似然函数 $L(x_1,x_2,\cdots,x_n;\theta)$达到最大的参数值$\hat{\theta}$，作为参数$\theta$的估计值.即取$\hat{\theta}$使

$$L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max_{\theta\in\Theta} L(x_1,x_2,\cdots,x_n;\theta)$$

这样得到的$\hat{\theta}$与样本值$x_1,x_2,\cdots,x_n$有关，常记为$\hat{\theta}(x_1,x_2,\cdots,x_n)$，称为参数$\theta$ 的**最大似然估计值**，而相应的统计量$\hat{\theta}(X_1,X_2,\cdots,X_n)$称为参数$\theta$的**最大似然估计量**.

若总体X属连续型，其概率密度$f(x;\theta),\theta\in\Theta$的形式已知，$\theta$为待估参数，$\Theta$是$\theta$ 可能取值的范围.设$X_1,X_2,\cdots,X_n$是来自X的样本，则$X_1,X_2,\cdots,X_n$的联合密度为

$$\prod_{i=1}^{n} f(x_i,\theta)$$

设$x_1,x_2,\cdots,x_n$是相应于样本$X_1,X_2,\cdots,X_n$的一个样本值，则随机点$(X_1,X_2,\cdots,X_n)$落在点$(x_1,x_2,\cdots,x_n)$的邻域（边长分别为 $dx_1,dx_2,\cdots,dx_n$的n维立方体）内的概率近似地为

$$\prod_{i=1}^{n} f(x_i;\theta)\,dx_i \tag{1}$$

其值随θ的取值而变化。与离散型的情况一样，取θ的估计值$\hat{\theta}$使概率(1)取到最大值，但因子$\prod_{i=1}^{n} dx_i$不随θ而变，故只考虑函数

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

的最大值.这里$L(\theta)$称为样本的**似然函数**.若

$$L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max_{\theta\in\Theta} L(x_1,x_2,\cdots,x_n;\theta)$$

则称$\hat{\theta}(x_1, x_2, \cdots, x_n)$为$\theta$的**最大似然估计值**，称$\hat{\theta}(X_1, X_2, \cdots, X_n)$为 θ的**最大似然估计量**。

这样，确定最大似然估计量的问题就归结为微分学中的求最大值的问题了。

在很多情况下，$p(x;\theta)$和$f(x;\theta)$关于θ可微，这时$\hat{\theta}$常可从方程

$$\frac{d}{d\theta}L(\theta) = 0$$

解得.又因$L(\theta)$和$\ln L(\theta)$在同一θ处取到极值，因此，θ的最大似然估计θ也可以从方程

$$\frac{d}{d\theta}\ln L(\theta) = 0 \tag{2}$$

求得，而从后一方程求解往往比较方便。(2)称为**对数似然方程**。

最大似然估计法也适用于分布中含多个未知参数$\theta_1, \theta_2, \cdots, \theta_k$的情况。这时，似然函数$L$是这些未知参数的函数. 分别令

$$\frac{\partial}{\partial \theta_i}L = 0, i = 1, 2, \cdots, k$$

或令

$$\frac{\partial}{\partial \theta_i}\ln L = 0, i = 1, 2, \cdots, k \tag{3}$$

解上述由k个方程组成的方程组，即可得到各未知参数$\theta_i(i = 1, 2, \cdots, k)$的最大似然估计值$\hat{\theta}_i$. (3)称为**对数似然方程组**。

定理 7.1 最大似然估计具有下述性质：设θ的函数$u = u(\theta), \theta \in \Theta$具有单值反函数$\theta = \theta(u), u \in \mathfrak{U}$。又假设 $\hat{\theta}$是X的概率分布中参数θ的最大似然估计，则$\hat{u} = u(\hat{\theta})$是$u(\theta)$的最大似然估计。这一性质称为最大似然估计的**不变性**.

事实上，因为$\hat{\theta}$是θ的最大似然估计，于是有

$$L(x_1, x_2, \cdots, x_n; \theta) = \max_{\theta \in \Theta} L(x_1, x_2, \cdots, x_n, \theta)$$

其中$x_1, x_2, \cdots, x_n$是X的一个样本值，考虑到$\hat{u} = u(\hat{\theta})$，且有$\hat{\theta} = \theta(\hat{u})$，上式可写成

$$L(x_1, x_2, \cdots, x_n; \theta(\hat{u})) = \max_{u \in \mathfrak{U}} L(x_1, x_2, \cdots, x_n; \theta(u))$$

这就证明了$\hat{u} = u(\hat{\theta})$是$u(\theta)$的最大似然估计。

当总体分布含有多个未知参数时，也具有上述性质。例如σ^2的最大似然估计为

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

函数$u = u(\sigma^2) = \sqrt{\sigma^2}$有单值反函数$\sigma^2 = u^2 (u \geq 0)$，根据上述性质，得到标准差$\sigma$的最大似然估计为

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

7.2 估计量的评选标准

定义 7.4 (无偏性) 若估计量$\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$的数学期望$E(\hat{\theta})$存在，且对于任意$\theta \in \Theta$有

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

则称$\hat{\theta}$是θ的**无偏估计量**。

定义 7.5 (有效性) 设$\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$与$\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$都是 θ的无偏估计量，若对于任意$\theta \in \Theta$，有

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

且至少对于某一个$\theta \in \Theta$上式中的不等号成立，则称$\hat{\theta}_1$较$\hat{\theta}_2$**有效**。

定义 7.6 (相合性) 设$\hat{\theta}(X_1, X_2, \cdots, X_n)$为参数$\theta$的估计量，若对于任意$\theta \in \Theta$，当$n \to \infty$时$\hat{\theta}(X_1, X_2, \cdots, X_n)$ 依概率收敛于θ，则称$\hat{\theta}$为θ的**相合估计量**。

即，若对任意$\theta \in \Theta$都满足：对任意$\varepsilon > 0$，有

$$\lim_{n \to \infty} P\{|\hat{\theta} - \theta| < \varepsilon\} = 1$$

则称$\hat{\theta}$是θ的**相合估计量**。

7.3 区间估计

定义 7.7 (置信区间) 设总体X的分布函数$F(x;\theta)$含有一个未知参数$\theta, \theta \in \Theta$（$\Theta$是$\theta$可能取值的范围），对于给定值$\alpha(0 < \alpha < 1)$，若由来自$X$的样本$X_1, X_2, \cdots, X_n$确定的两个统计量$\underline{\theta} = \underline{\theta}(X_1, X_2, \cdots, X_n)$ 和$\overline{\theta} = \overline{\theta}(X_1, X_2, \cdots, X_n)(\theta < \overline{\theta})$，对于任意$\theta \in \Theta$满足

$$P\{\underline{\theta}(X_1, X_2, \cdots, X_n) < \theta < \overline{\theta}(X_1, X_2, \cdots, X_n)\} \geq 1 - \alpha$$

则称随机区间$(\underline{\theta}, \overline{\theta})$是$\theta$的置信水平为$1-\alpha$的**置信区间**，$\underline{\theta}$和$\overline{\theta}$分别称为置信水平为$1-\alpha$的双侧置信区间的**置信下限**和**置信上限**，$1-\alpha$称为**置信水平**。

当X时离散型随机变量时，对于给定的α，常常找不到区间$(\underline{\theta}, \overline{\theta})$使得$P\{\underline{\theta} < \theta < \overline{\theta}\}$恰为$1-\alpha$。此时去找区间$(\underline{\theta}, \overline{\theta})$使得$P\{\underline{\theta} < \theta < \overline{\theta}\}$至少为$1-\alpha$，且尽可能地接近$1-\alpha$.

定理 7.2 寻求未知参数θ的置信区间的具体做法：

1° 寻求一个样本$X_1, X_2, \cdots, X_n$和θ的函数$W = W(X_1, X_2, \cdots, X_n; \theta)$，使得$W$的分布不依赖于$\theta$以及其他未知参数，称具有这种性质的函数$W$为**枢轴量**.

2° 对于给定的置信水平$1-\alpha$，定出两个常数a, b，使得

$$P\{a < W(X_1, X_2, \cdots, X_n; \theta) < b\} = 1 - \alpha$$

若能从$a < W(X_1, X_2, \cdots, X_n; \theta) < b$得到与之等价的$\theta$的不等式$\underline{\theta} < \theta < \overline{\theta}$，其中 $\underline{\theta} = \underline{\theta}(X_1, X_2, \cdots, X_n), \overline{\theta} = \overline{\theta}(X_1, X_2, \cdots, X_n)$都是统计量。那么$(\underline{\theta}, \overline{\theta})$就是$\theta$的一个置信水平为$1-\alpha$的置信区间。

枢轴量$W(X_1, X_2, \cdots, X_n; \theta)$的构造，通常可以从$\theta$的点估计着手考虑。常用的正态总体的参数的置信区间可以用上述步骤推得。

7.4 正态总体均值与方差的区间估计

7.5 （0-1）分布参数的区间估计

7.6 单侧置信区间

8 假设检验

8.1 假设检验

9　方差分析及回归分析

9.1　单因素试验的方差分析