第二部分
概率分布的基本性质

- 本章内容：

  随机变量的概率分布函数的基本性质：平均值、方差、协方差矩阵、矩、…

概率密度函数（Probability Density Function)

定义：

X：连续型随机变量；
Ω：样本空间（X的值域）
X的值落入区间[*x*,*x*+*dx*]的概率：

$$p(x \le X \le x + dx) = f(x)dx$$

其中：*f(x)*被称为随机变量X的概率密度函数（p.d.f），表示单位长度下的概率。
归一化条件（normalization condition）：

$$\int_{\Omega} f(x)dx = 1$$ 表示：在样本空间内，随机变量X总会取某一值

性质：

1. 对所有的*x*值，$f(x) \ge 0$
2. *f(x)*是单值函数
3. *f(x)*是非奇异的

简称分布函数

定义：

$$F(x)=\int_{x_{\min}}^{x} f(x')dx'$$

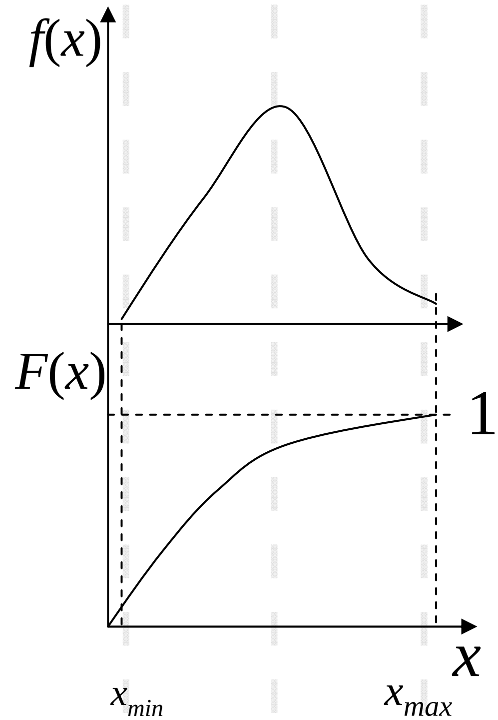
其中：x_{min}是随机变量X的取值下限

意义：

表示随机变量X的取值小于某一值x的概率，即

$$F(x)=p(X\le x),\qquad x_{\min}\le x\le x_{\max}$$

离散型随机变量可以定义累积分布函数



性质：

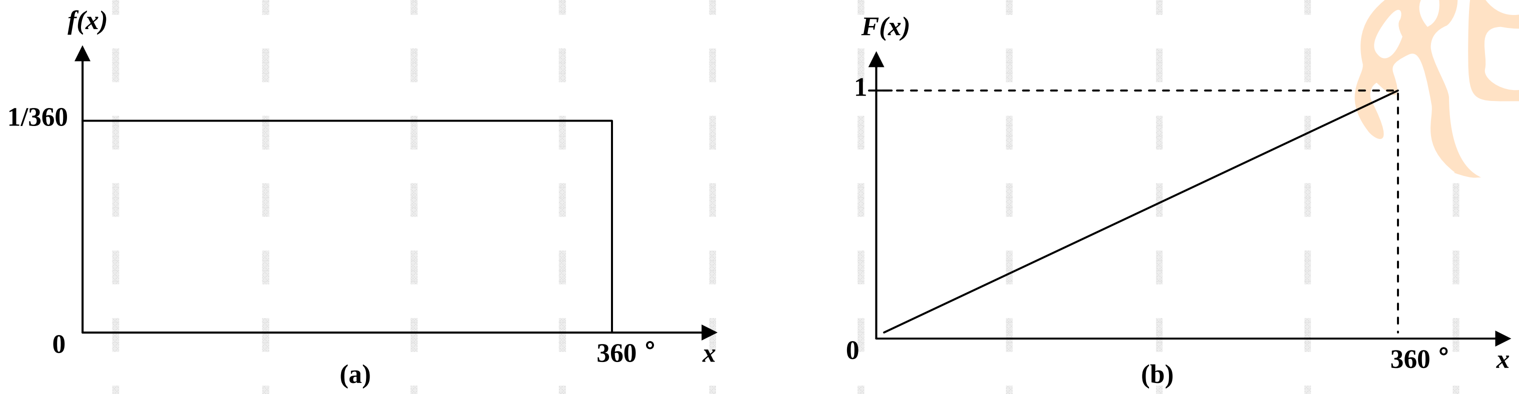
1、$0\le F(x)\le 1,\qquad x_{\min}\le x\le x_{\max}$

2、$F(x_{min})=0$，$F(x_{max})=1$

3、若$x_1<x_2$，则$F(x_1)<F(x_2)$，即$F(x)$是单调升函数

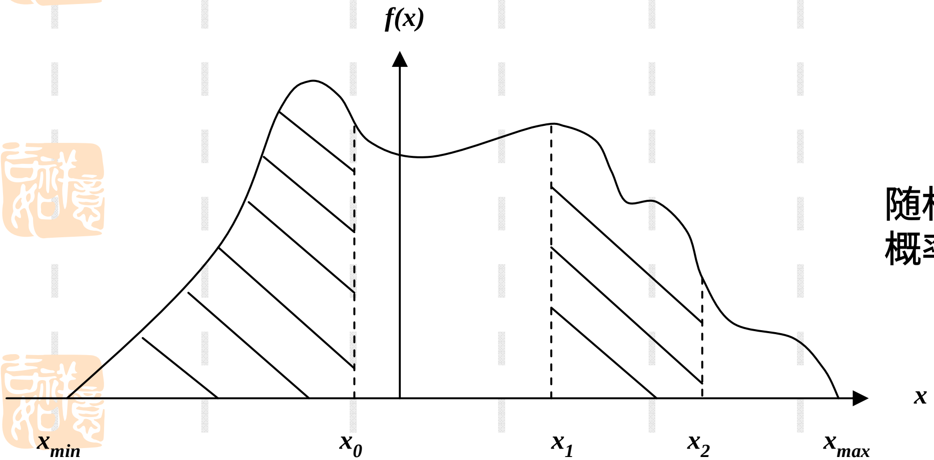
4、$p(x_1\le x\le x_2)=F(x_2)-F(x_1)$

5、$F(x_+)-F(x_-)=\lim\limits_{n\to\infty}\int_{x-\frac{1}{n}}^{x+\frac{1}{n}} p(x')dx'=0$，即x取特定值的几率为0。



一个均匀分布的例子

时钟角度X的p.d.f.— $f(x)$（a）和分布函数$F(x)$（b）



随机变量X的取值x在区间$[x_1, x_2]$的概率是概率密度函数$f(x)$曲线下相应区间的面积

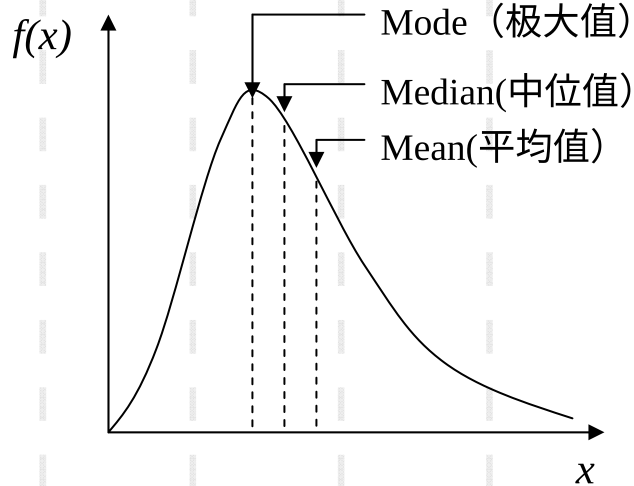
概率密度函数的性质
(Properties of the probability)

f(x)包含了随机变量X的所有信息，其性质确定了X的分布

最可几值(mode)：使$f(x)$取极大的x值，x_p

中位值 (median)：$F(x_{median})=1/2$

平均值(mean)：$\bar{x} \equiv \int_\Omega x f(x) dx$



统计物理中，麦克斯韦速度分布律给出：

$$v_p : \bar{v} = 1 : \frac{2}{\sqrt{\pi}} = 1 : 1.128$$

一、函数的期望值（Expectation)

定义：

$f(x)$：随机变量X的概率密度函数
$g(x)$：随机变量X的函数
$g(x)$的期望值（对$g(x)$的加权平均值）：

$$E[g(x)] \equiv \int_{\Omega} g(x) f(x) dx$$

E[$g(x)$]是一个常数，与x无关，是函数$g(x)$的平均值或中值的一个量度

$f(x)$对应于量子力学或统计物理中的态密度（如麦克斯韦分布、玻色－爱因斯坦分布、费米－狄拉克分布）

$g(x)$可以理解为一个物理量算符

$$1 = \int_{\Omega} f(x) dx = \int_{\Omega} < \Phi(x) | \Phi(x) > dx$$

$$E[g(x)] = \int_{\Omega} g(x) f(x) dx = \int_{\Omega} < \Phi(x) | \hat{g}(x) | \Phi(x) > dx$$

一、函数的期望值（Expectation)

定义：

积分学第一中值定理（Mean Value Theorem）

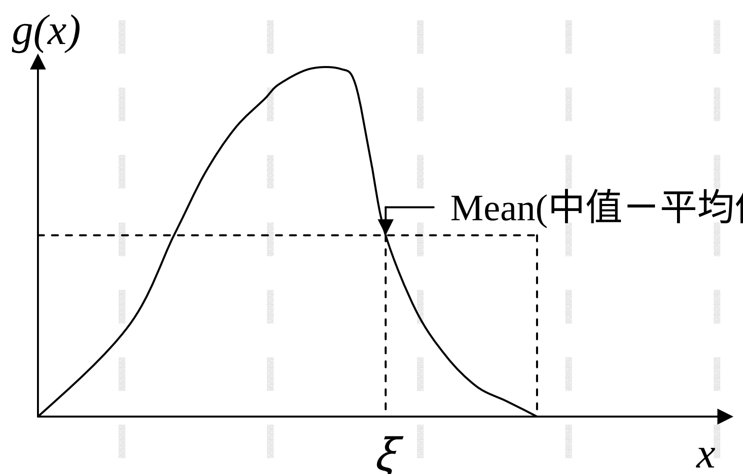
对连续函数$g(x)$，在区间[a, b]上存在 ξ，使得

$$\int_a^b g(x)dx = g(\xi)(b-a)$$

$g(\xi)$称为$g(x)$的积分中值，或平均值。实际上就是算术平均值，对离散的函数g，就很容易看出来。

$$\sum_{i=1}^{n} g_i = \bar{g} \cdot n$$

这意味着可以找到一个点ξ，使得g(x)下的面积等价于一个矩形面积，但这不是统计学中通常定义的平均值（见下面）。



一、函数的期望值（Expectation)

定义：

积分学第二中值定理

对连续函数$f(x)$和$g(x)$，在区间[a, b]上存在 ξ，使得

$$\int_a^b g(x)f(x)dx = f(a)\int_a^{\xi} g(x)dx + f(b)\int_{\xi}^b g(x)dx$$

$g(\xi)$称为$g(x)$的积分中值（加权平均值），$f(x)$为权函数。

第一中值定理是第二中值定理在$f(x)=1$时的特例。

平均值可以指x^2的、x^3的等等任意函数的平均值。

概率论中的平均值特指自变量x的加权平均值，即$g(x)=x$

如果把求平均值的运算E看作一个算符，它具有性质：

若*a*是常数，则

$$E(a)=a$$

$$E[ag(x)]=aE[g(x)]$$

$$E[a_1g_1(x)+a_2g_2(x)]=a_1E[g_1(x)]+a_2E[g_2(x)]$$

即，E是线性算符

函数的方差（Variance）

定义：

$$V[g(x)]\equiv E\{g(x)-E[g(x)]\}^2$$

$$\equiv\int_\Omega\{g(x)-E[g(x)]\}^2 f(x)dx$$

意义：

g(x)在其期望值周围的离散程度

二、随机变量的平均值和方差（Mean Value and Variance）

如取g(x)=x，则得随机变量X的平均值和方差

平均值： $\mu \equiv E(x) \equiv \int_{\Omega} x f(x) dx$

方差： $\sigma^2 \equiv V(x) \equiv E(x-\mu)^2 \equiv \int_{\Omega} (x-\mu)^2 f(x) dx$

σ：随机变量X的标准偏差（**Standard Deviation**）

方差（**Variance，Dispersion**）物理意义：

随机变量概率密度函数f(x)在期望值周围的离散程度，亦即由于随机的统计性所造成的随极变量的取值在期望值附近的起伏的大小。

平均值与方差之间的关系：

$$\sigma^2=E(x-\mu)^2=E(x^2-2\mu x+\mu^2)=E(x^2)-\mu^2=E(x^2)-[E(x)]^2$$

$$\sigma^2=E(x^2)-[E(x)]^2$$

实验上常把物理量的测量结果表示成：$\mu \pm \sigma$

三、矩（moment）

定义：

对正整数k(k=1,2,…)，

$$\mu'_k \equiv E(x^k) = \int_\Omega x^k f(x)dx$$ ←k阶原点矩

$$\mu_k \equiv E(x-\mu)^2 = \int_\Omega (x-\mu)^k f(x)dx$$ ←k阶中心矩

分别称为随机变量X的k阶原点矩和中心矩。

显然，随机变量X的平均值和方差分别为

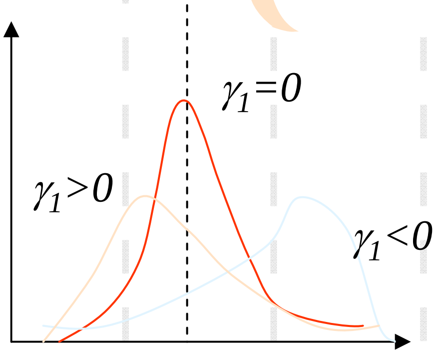
$$\mu = \mu'_1$$ ←1阶原点矩

$$\sigma^2 = \mu_2$$ ←2阶中心矩

高阶矩用于研究在$|x-\mu|$较大时$f(x)$的特性：

1. 非对称系数，或偏度（**Skewness):**

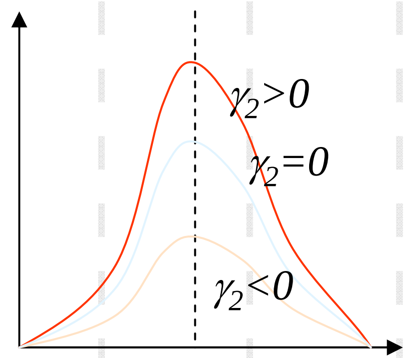
表征$f(x)$对平均值的不对称程度



$$\gamma_1 \equiv \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{E(x-\mu)^3}{\sigma^3}$$

2. 峰度系数（**Peakedness):**

$f(x)$曲线的尖锐程度（与正态分布相比）



$$\gamma_2 \equiv \frac{\mu_4}{(\mu_2)^2} - 3 = \frac{E(x-\mu)^4}{\sigma^4} - 3$$

均匀分布的 峰度系数 $\gamma_2 = -1.2$

定义：

复值随机变量e^{itx}的期望值

$$\Phi(t) \equiv E(e^{itx}) = \int_{-\infty}^{\infty} e^{itx} f(x)dx$$

称为随机变量X的特征函数。

$\Phi(t)$作Fourier展开，可以表示成各阶原点矩为系数的级数和。

$\Phi(t)$与$f(x)$构成Fourier变换对，即

$$f(x) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\Phi(t)dt$$

由特征函数可很容易地求出随机变量X的各阶矩

原点矩： $$\mu'_k = \frac{\partial^k \Phi(t)}{\partial^k (it)}\Big|_{t=0}, \qquad k = 1,2,\ldots$$

中心矩： $$\mu_k = \frac{\partial^k \Phi_\mu(t)}{\partial^k (it)}\Big|_{t=0}, \qquad k = 1,2,..$$

$$\Phi_\mu(t) = \int_{-\infty}^{\infty} e^{it(x-\mu)} f(x)dx = E[e^{it(x-\mu)}]$$

特征函数的性质：

1. $\Phi(0)=1, \qquad |\Phi(t)|\le 1$

2. 如果a和b为常数，则ax+b的特征函数为

$$\Phi_{ax+b}(t)=e^{ibt}\Phi_x(at)$$

多个随机变量的分布

上述单个随机变量的概率分布的特征可推广到多个随机变量的情况

一、联合概率密度函数(joint probability density function)

N维随机向量：

n个随机变量的整体，$\vec{x} = (x_1, x_2, \ldots x_n)$

联合分布函数：

$$F(x_1, x_2, \ldots x_n) = p(X_1 \le x_1, X_2 \le x_2, \ldots X_n \le x_n)$$
$$= \int_\Omega f(\vec{x}) d\vec{x}$$

联合概率密度函数： $f(\vec{x})$

$f(\vec{x})$ 的性质：

在**n**维空间中是非负的、单值函数，且满足归一化条件：

$$\int_\Omega f(x_1, x_2, \ldots x_n) dx_1 dx_2 \ldots dx_n = \int_\Omega f(\vec{x}) d\vec{x} = 1$$

二、期望值和方差

设 $g(\vec{x}) = g(x_1, x_2, \dots x_n)$ 是随机向量 $\vec{x}$ 的函数，则 $g(\vec{x})$ 的期望值和方差定义为

$$E[g(\vec{x})] \equiv \int_\Omega g(\vec{x}) f(\vec{x}) d\vec{x}$$

$$V[g(\vec{x})] \equiv E[g(\vec{x}) - E[g(\vec{x})]]^2$$

$$= \int_\Omega (g(\vec{x}) - E[g(\vec{x})])^2 f(\vec{x}) d\vec{x}$$

三、协方差矩阵和相关系数

(Covariance Matrix, Correlation Coefficient)

随机向量 $\vec{x}$ 的某一分量 x_i 的平均值和方差的定义为：

$$\mu_i \equiv E(x_i) = \int_\Omega x_i f(\vec{x}) d\vec{x}$$

$$\sigma^2 \equiv E(x_i - \mu_i)^2 = \int_\Omega (x_i - \mu_i)^2 f(\vec{x}) d\vec{x}$$

各分量的方差表示各分量偏离其期望值的分散程度，为表示各分量间的相互关联的数字特征，引入协方差矩阵和相关系数的概念

协方差矩阵元的定义：

$$V_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)] = \int_\Omega (x_i - \mu_i)(x_j - \mu_j) f(\vec{x}) d\vec{x}$$
$$= E(x_i x_j - \mu_i x_j - x_i \mu_j + \mu_i \mu_j) = E(x_i x_j) - \mu_i E(x_j) - E(x_i)\mu_j + \mu_i \mu_j$$
$$= E(x_i x_j) - E(x_i)E(x_j)$$

其中，μ_i 和 μ_j 分别为变量 x_i 和 x_j 的平均值。

协方差矩阵的特征：

（1）$V(x)$ 是对称矩阵；

（2）对角元素是随机向量各分量的方差；

$$\sigma^2 \equiv V_{ij} = E(x_i^2) - [E(x_i)]^2$$

（3）非对角元素称为变量 x_i 和 x_j 的协方差，记为：$cov(x_i,x_j)$

$$V_{ij}(i \neq j) = \mathrm{cov}(x_i, x_j) \equiv V_{ij} = E(x_i x_j) - E(x_i)E(x_j)$$

协方差可正、可负。

相关系数的定义：

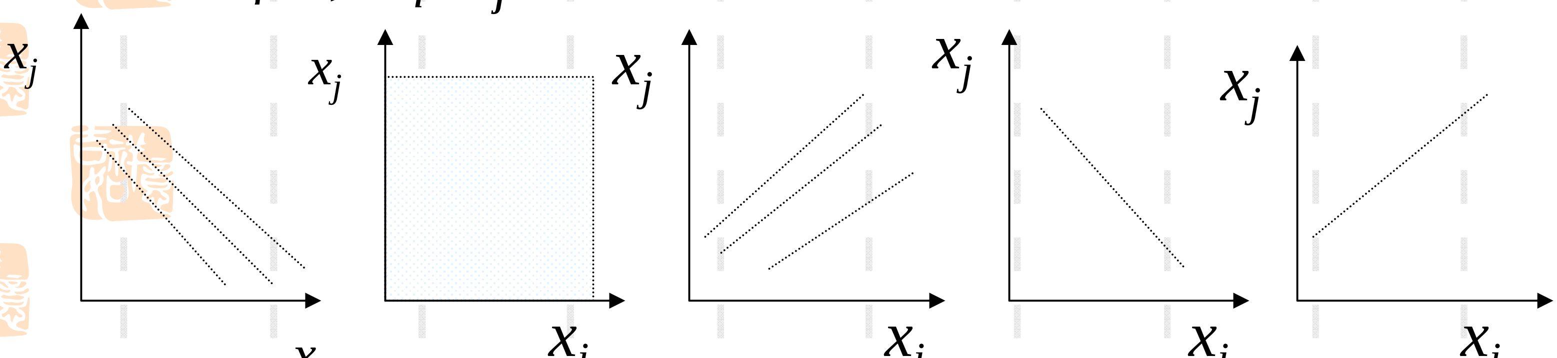
$$\rho(x_i, x_j) = \frac{V_{ij}}{\sqrt{V_{ii}V_{jj}}} = \frac{\mathrm{cov}(x_i, x_j)}{\sigma_i \sigma_i}$$

可证：

$$-1 \le \rho(x_i, x_j) \le 1$$

相关系数的含义：

- 如果ρ=+1（或-1）,则称x_i和x_j为完全正相关(或负相关);
- 如果ρ=0, 则称x_i和x_j不相关;
- 如果ρ>0,则x_i和x_j的变化趋势相同
- 如果ρ<0,则x_i和x_j的变化趋势相反



四、独立变量(Independent variables)

如果随机向量 $\vec{x}=(x_1,x_2,\cdots x_n)$ 的联合概率密度函数可写为

$$f(x_1,x_2\cdots,x_n)=f_1(x_1)\cdot f_2(x_2)\cdots f_n(x_n)$$

则称随机变量$x_1,x_2,\ldots,x_n$是相互独立的

对于相互独立的随机变量，它们的协方差和相关系数为零，即不相关；反之，不成立。

$$\mathrm{cov}(x_i,x_j)=0,\rho(x_i,x_j)=0$$

$$\begin{aligned}E(x_ix_j)&=\int_\Omega x_ix_jf(x_i,x_j)dx_idx_j\\&=\int_\Omega x_ix_jf(x_i)f(x_j)dx_idx_j\\&=\int_\Omega x_if(x_i)dx_i\int_\Omega x_jf(x_j)dx_j\\&=E(x_i)E(x_j)\end{aligned}$$

$$\mathrm{cov}(x_i,x_j)=E(x_ix_j)-E(x_i)E(x_j)=0$$

五、边缘概率密度函数，条件概率密度函数（Marginal and Conditional p.d.f)

边缘概率密度函数

概率密度函数 $f(\vec{x})$在某一子空间上的投影称为随机向量 $\vec{x}$ 对这一子空间的边缘概率密度函数

例：随机向量 $\vec{x}$ 对分量x_1的边缘密度函数定义为：

$$h_1(x_1) \equiv \int_{x_{2\min}}^{x_{2\max}} \cdots \int_{x_{n\min}}^{x_{n\max}} f(x_1, x_2, \cdots x_n) dx_2 \cdots dx_n$$

如果$x_1, x_2, \ldots x_n$是相互独立的，则有

$$h_1 = f_1(x_1), \quad h_2 = f_2(x_2), \ldots.$$

即：相互独立的随机变量的联合概率密度函数可因式分解为对各分量的边缘概率密度函数之积

条件概率密度函数

$$f(x_2, x_3, \cdots x_n \mid x_1) \equiv \frac{f(x_1, x_2, \cdots x_n)}{h_1(x_1)}$$

六、联合特征函数

随机向量 $\vec{X}$ 的联合特征函数的定义：

$$\Phi(t_1,t_2,\cdots t_n)\equiv E(e^{it_1x_1+it_1x_1+\cdots+it_1x_n})$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}e^{it_1x_1+it_1x_1+\cdots+it_1x_n}f(x_1,x_2,\cdots,x_n)dx_1dx_2\cdots dx_n$$

如果$x_1,x_2,\ldots x_n$是相互独立的，则有

$$\Phi(t_1,t_2,\cdots t_n)=\Phi(t_1)\cdot\Phi(t_2)\cdots\Phi(t_n)$$

其中，$\Phi(t_i)$为随机变量x_i的特征函数

证明：考虑两个变量的情况：如果x_1和x_2为相互独立的随机变量，则有

$$f(x_1,x_2)=f_1(x_1)\cdot f_2(x_2)$$

$$\Phi(t_1,t_2)=E(e^{it_1x_1}\cdot e^{it_2x_2})=E(e^{it_1x_1})\cdot E(e^{it_2x_2})=\Phi(t_1)\cdot\Phi(t_2)$$

由联合特征函数可得到：

$$E(x_1^r x_2^s)=\frac{\partial^{r+s}\Phi}{\partial(it_1)^r\partial(it_2)^s}\Big|_{t=0}$$

随机变量的线性函数

设$g(x_1,x_2,\ldots,x_n)$是n个随机变量的线性函数：

$$g(x_1,x_2,\cdots,x_n)=\sum_{i=1}^{n}a_ix_i$$

求$g(x_1,x_2,\ldots,x_n)$的期望值和方差

1、期望值： $$E[\sum_{i=1}^{n}a_ix_i]=\sum_{i=1}^{n}E[a_ix_i]=\sum_{i=1}^{n}a_iE[x_i]=\sum_{i=1}^{n}a_i\mu_i$$

2、方差： $$V[\sum_{i=1}^{n}a_ix_i]=E\{\sum_{i=1}^{n}a_ix_i-E[\sum_{i=1}^{n}a_ix_i]\}^2$$

$$=E\{\sum_{i=1}^{n}a_ix_i-\sum_{i=1}^{n}a_i\mu_i\}^2=E\{\sum_{i=1}^{n}a_i(x_i-\mu_i)\}^2$$

$$=E\{\sum_{i=1}^{n}a_i^2(x_i-\mu_i)^2+\sum_{i\neq j}a_ia_j(x_i-\mu_i)(x_j-\mu_j)\}$$

$$=\sum_{i=1}^{n}a_i^2E(x_i-\mu_i)^2+\sum_{i\neq j}a_ia_jE[(x_i-\mu_i)(x_j-\mu_j)]$$

$$=\sum_{i=1}^{n}a_i^2V(x_i)+\sum_{i\neq j}a_ia_j\,\mathrm{cov}(x_i,x_j)$$

$$\therefore \qquad V[\sum_{i=1}^{n} a_i x_i] = \sum_{i=1}^{n} a_i V_{ii} + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_i a_j V_{ij}$$

若$x_1,x_2,\ldots,x_n$是相互独立的随机变量，$V_{ij}=0$，则

$$V[\sum_{i=1}^{n} a_i x_i] = \sum_{i=1}^{n} a_i V_{ii}$$

例：n个相互独立的随机变量$x_1,x_2,\ldots,x_n$,具有相同的平均值μ和方差σ^2:

定义算术平均值:

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

求$\bar{x}$的期望值和方差

$$a_i = \frac{1}{n}, \mu_i = \mu, V_{ii} = \sigma_i^2 = \sigma^2$$

$$E(\bar{x}) = \sum_{i=1}^{n} a_i \mu_i = \sum_{i=1}^{n} \frac{1}{n}\mu = \mu$$

$$V(\bar{x}) = \sum_{i=1}^{n} a_i^2 \sigma_i^2 = \sum (\frac{1}{n})^2 \sigma^2 = \frac{\sigma^2}{n}$$

变量变换
(Change of variables)

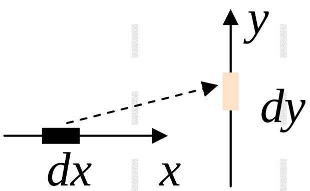
X: 连续型的随机变量, PDF: $f(x)$

$y = y(x)$：x的函数, 也是随机变量. 求$y(x)$的概率密度函数$g(x)$

1、若随机变量x和y是一一对应的:

$$[x, x+dx] \rightarrow [y, y+dy]$$

X的取值在$[x, x+dx]$的概率==Y的取值在$[y, y+dy]$的概率：



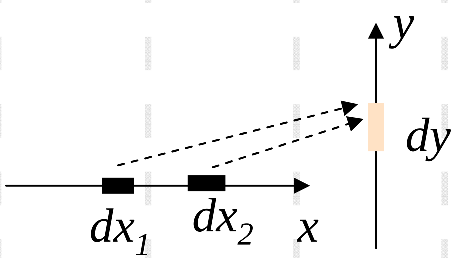
$$f(x)dx=g(y)dy \rightarrow \quad g(y) = f(x)\left|\frac{dx}{dy}\right|$$

取绝对值是为了保证g(y)是非负的

2、若随机变量x和y不是一一对应的:

即有n个区间$[x,x+dx] \rightarrow [y,y+dy]$,

需要对这n个区间求和



$$g(y) = \sum f(x)\left|\frac{dx}{dy}\right|$$

3、推广到n个随机变量的情况:

$$\vec{x}=(x_1,x_2,\cdots x_n)\rightarrow\vec{y}=(y_1,y_2,\cdots y_n)$$

$$g(\vec{y})=f(\vec{x})\left|J\right|$$

$$\left|J\right|=\left|\frac{\partial(x_1,x_2,\cdots x_n)}{\partial(y_1,y_2,\cdots y_n)}\right|=\begin{vmatrix}\frac{\partial x_1}{\partial y_1},\frac{\partial x_1}{\partial y_2}\cdots\frac{\partial x_1}{\partial y_n}\\ \vdots \\ \frac{\partial x_n}{\partial y_1},\frac{\partial x_n}{\partial y_2}\cdots\frac{\partial x_n}{\partial y_n}\end{vmatrix}$$

➔**Jacobian行列式**

误差传播
(Propagation of errors)

实验测量的物理量可分为:

直接测量量:其值是用实验仪器直接测量的

间接测量量:其值是用直接测量量的结果是通过适当的公式推断出来的

如何通过直接测量量的误差推导出间接测量量的误差?→误差传播公式

一、单一函数的情况

设y是随机变量(直接测量量) $\vec{x}=(x_1,x_2,\cdots x_n)$的函数:

$$y = y(x_1, x_2, \cdots x_n) = y(\vec{x})$$

$\vec{x}$

协方差矩阵 $V(\vec{x})$

平均值: $\vec{\mu} = (\mu_1, \mu_2, \cdots \mu_n)$

求y的方差

$$y(\vec{x}) \approx y(\vec{\mu}) + \sum_{i=1}^{n} (x_i - \mu_i) \tfrac{\partial y}{\partial x_i} \Big|_{\vec{x}=\vec{\mu}}$$

$$E[y(\vec{x})] \approx y(\vec{\mu})$$

$$\therefore V[y(\vec{x})] = E\{y(\vec{x}) - E[y(\vec{x})]\}^2 \approx E[y(\vec{x}) - y(\vec{\mu})]^2$$

$$y(\vec{x}) - y(\vec{\mu}) \approx \sum_{i=1}^{n} (x_i - \mu_i) \frac{\partial y}{\partial x} \Big|_{\vec{x}=\vec{\mu}}$$

$$\therefore V[y(\vec{x})] \approx \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial y}{\partial x_i} \Big|_{\vec{x}=\vec{\mu}} \frac{\partial y}{\partial x_j} \Big|_{\vec{x}=\vec{\mu}} E[(x_i - \mu_i)(x_j - \mu_j)]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial y}{\partial x_i} \Big|_{\vec{x}=\vec{\mu}} \frac{\partial y}{\partial x_j} \Big|_{\vec{x}=\vec{\mu}} V_{ij}$$

误差传播定律：

$$V[y(\vec{x})] \approx \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial y}{\partial x_i}\Big|_{\vec{x}=\vec{\mu}} \frac{\partial y}{\partial x_j}\Big|_{\vec{x}=\vec{\mu}} V_{ij}$$

如果$x_1,x_2,\ldots,x_n$是相互独立的，则有

$$V[y(\vec{x})] \approx \sum_{i=1}^{n}\left(\frac{\partial y}{\partial x_i}\Big|_{\vec{x}=\vec{\mu}}\right)^2 V_{ii}$$

将方差用σ^2代替，则得

$$\sigma^2(y) \approx \sum_{i=1}^{n}\left(\frac{\partial y}{\partial x_i}\Big|_{\vec{x}=\vec{\mu}}\right)^2 \sigma_i^2(x) + 2\sum_{i<j,\, j=2}^{n}\left(\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right)_{\vec{x}=\vec{\mu}} \mathrm{cov}(x_i, x_j)$$

例：算术平均值的方差

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

$$\frac{\partial y}{\partial x_i} = \frac{1}{n}, \qquad \sigma_i(x) = \sigma$$

$$\therefore\ \sigma^2(\bar{x}) = \sum_{i=1}^{n}\left(\frac{\partial y}{\partial x_i}\right)^2 \sigma^2 = \frac{\sigma^2}{n}$$

二、多个函数的情况，矩阵表示

设有一组m个函数$y_1,y_2,\ldots y_m$都依赖于n维随几向量$(x_1,x_2,\ldots x_n)$

$$y_k=y_k(x_1,x_2,\ldots x_n)=y_k(\underline{x}),\ k=1,2,\ldots m$$

$$y_k(\vec{x})=y_k(\vec{\mu})+\sum_{i=1}^{n}(x_i-\mu_i)\frac{\partial y_k}{\partial x_i}\Big|_{\vec{x}=\vec{\mu}}+\cdots,\qquad k=1,2,\cdots m$$

$$E[y_k(\vec{x})]\approx y_k(\vec{\mu}),\qquad k=1,2,\cdots m$$

$$\begin{aligned}V_{kl}(\vec{y})&=E\{[y_k(\vec{x})-E(y_k(\vec{x}))][y_l(\vec{x})-E(y_l(\vec{x}))]\}\\&=E\{[y_k(\vec{x})-y_k(\vec{\mu})][y_l(\vec{x})-y_l(\vec{\mu})]\}\\&\approx\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial y_k}{\partial x_i}\frac{\partial y_l}{\partial x_j}E[(x_i-\mu_i)(x_j-\mu_j)]\\&=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial y_k}{\partial x_i}\frac{\partial y_l}{\partial x_j}V_{ij}(\vec{x})\end{aligned}$$

$\Leftarrow$ 随机向量$y_1,y_2,\ldots y_m$的协方差矩阵元

→误差传播定律的一般形式

变量$y_1, y_2, \dots y_m$的误差为对角元素的平方根

$$\sigma_k^2 = V_{kk}(\vec{y}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial y_k}{\partial x_i}\frac{\partial y_k}{\partial x_j}V_{ij}(\vec{x})$$

➔依赖于随即向量X的协方差项

如果x_i是相互独立的，

$$V_{ij}(\vec{x}) = 0, \text{当} i \neq j \text{时}$$

$$\sigma_k^2 = V_{kk}(\vec{y}) = \sum_{i=1}^{n}\left(\frac{\partial y_k}{\partial x_i}\right)^2 V_{ii}(\vec{x}) = \sum_{i=1}^{n}\left(\frac{\partial y_k}{\partial x_i}\right)^2 \sigma_i^2(\vec{x})$$

矩阵表示：

$$V(\vec{y}) = SV(\vec{x})S^T$$

$$S_{ki} = \frac{\partial y_k}{\partial x_i}\Big|_{\vec{x}=\vec{\mu}}$$

分离型随机变量的概率分布

概率分布：一组概率值p_r表示，p_r满足归一化条件：

$$\sum_r p_r = 1$$

p_r :分离型随即变量取值为r的概率

期望值和方差的定义：与连续型的随机变量类似，积分→求和

$$E(r) = \sum_r r p_r$$

$$V(r) = \sum_r (r - E(r))^2 p_r = E(r^2) - [E(r)]^2$$

概率产生函数（**probability generating function)** ←特征函数

$$G(z) \equiv E(z^r) = \sum_r z^r p_r$$

利用该函数可计算变量r的各阶矩：

$$\because G'(z) = \sum_r r z^{r-1} p_r, \qquad G''(z) = \sum_r r(r-1) z^{r-2} p_r$$

$$\therefore G'(1) = \sum_r r p_r = E(r), \qquad G''(1) = \sum_r r(r-1) p_r = \sum_r r^2 p_r - \sum_r r p_r = E(r^2) - E(r)$$

$$\therefore E(r) = G'(1), \qquad V(r) = G''(1) + G'(1) - (G'(1))^2$$

样本
(Sampling)

一、总体和样本（universe and sample)

总体（或母体）：

研究对象的所有可能的观测结果

- 在物理实验中，总是用一些随机变量来描述某一物理系统，这些变量的概率密度函数描述了总体的特征
- 如果能在相同的条件下对描述物理系统的随机变量进行无限多次的测量，则可用概率密度函数来概括所有可能的实验结果；

样本（sample)：

在实际实验中，测量的次数总是有限的，若实验的次数为n，对某个物理量的测量值为$x_1, x_2, \ldots x_n$，则称这组测量值构成了容量为n的样本。

- 样本只是总体的一个子集，希望能从该样本推断出总体的特征;
- 样本是随机的：不同的样本对总体的特性的推断有差异,但基本类似

二、样本的特性

希望用实验样本推断出所研究的总体的特性

样本平均值： $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

样本方差： $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

➔样本相对于其平均值的离散程度

$\bar{x}$和s^2是随机变量$\mathbf{x_i}$的函数➔也是随机变量

- 如果从总体中抽取几组容量都为**n**的样本，每组样本的平均值和方差将是不同的；
- 样本平均值和方差将具有自己的分布，其分布依赖于总体的分布和样本的容量；

特例：总体满足正态分布，则样本平均值和方差具有以下的性质：

1. 样本平均值和方差是相互独立的随机变量；
2. 样本平均值服从正态分布；
3. 样本方差服从χ^2分布

三、由样本得出的推论

实验的目的就是要用有限的样本的特性推断出总体的特性，希望样本能在某些方面代表所研究的总体的特性。

样本平均值和方差可用于估计总体的平均值μ和方差σ^2

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

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

当n很大时，样本的特性应趋近于总体的特性

$$\hat{\mu} = \bar{x} \to \mu$$

$$\hat{\sigma}^2 = s^2 \to \sigma^2$$

➔广义的收敛定律

直观理解：

1、样本平均值的期望值和方差：

$$E(\bar{x}) = \mu, V(\bar{x}) = \sigma^2 / n$$

因此，当n很大时，**V(x)➔0**

2、s^2的期望值和方差

$$E(s^2) = \sigma^2 = \mu_2, V(s^2) = \tfrac{1}{n}(\mu_4 - \mu_2^2) + \tfrac{2}{(n-1)n}\mu_2^2$$

$$\therefore \text{当} n \to \infty \text{时，} V(s^2) \to 0$$

因此,如果选择容量足够大的样本,则对总体参数的估计值可达到要求的精度

➔估计式的一致性(**consistency**)

四、大数定理(Law of Large Numbers)

当$n \to \infty$时，$\hat{\mu} = \bar{x} \to \mu$ 是大数定理的一个结果

大数定理:

设$x_1, x_2, \ldots$是一组具有相同分布的独立的随机变量（平均值都为μ）,对于其中的前n个变量，定义算术平均值为

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

给定任意的两个正整数ε和δ,存在着正整数N,使得当 $n \geq N$ 时, $|\bar{x} - \mu| > \varepsilon$的概率小于$\delta$

$$P(|\bar{x} - \mu| > \varepsilon) < \delta$$

➔给出了当n很大时,算术平均值的行为

大数定理与随机变量的方差没有关系,即使方差不存在,该定理也成立

如果x_i的方差存在，利用切比雪夫不等式

$$P(|\bar{x}-\mu|\geq\varepsilon)\leq\frac{V(\bar{x})}{\varepsilon^2}=\frac{\sigma^2}{n\varepsilon^2}$$

当$n\to\infty$时，$|\bar{x}-\mu|\geq\varepsilon$ 的概率可以任意小