第一章 概率论的基本概念

考点 1：随机试验

定义：

1．可以在相同的条件下重复地进行.

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一个结果会出现.

在概率论中,我们将具有上述三个特点的试验称为随机试验。

考点 2：样本空间

定义：将随机试验 E 的所有可能结果组成的集合称为 E 的样本空间,记为 S；样本空间的元素,即 E 的每个结果,称为样本点。

补充：对于随机试验,尽管在每次试验之前不能预知试验的结果,但试验的所有可能结果组成的集合是已知的。

考点 3：随机事件

定义：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.

补充：由一个样本点组成的单点集,称为基本事件。

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发生,$\emptyset$称为不可能事件。

考点 4：随机事件间的关系

设试验E的样本空间为S，而A、B、$A_k(k=1,2,\dots)$是S的子集.

定义：

包含关系：若$A\subset B$,则称事件B包含事件A,这指的是事件A发生必导致事件B发生.

相等关系：若$A\subset B$且$A\supset B$,即$A=B$，则称事件A与事件B相等.

互不相容：若事件$A\cap B=\emptyset$，则称事件 A 与 B 是互不相容的,或互斥的.这指的是事件 A 与事件 B 不能同时发生.基本事件是两两互不相容的.

对立关系：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$\bar{A}$.$\bar{A}=S-A$.

考点 5：随机事件之间的运算

定义：

和事件：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$\bigcup_{k=1}^{n}A_k$为n个事件$A_1,A_2,\dots,A_n$的和事件；称$\bigcup_{k=1}^{\infty}A_k$为可列个事件$A_1,A_2,\dots$的和事件。

积事件：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. $A\cap B$也记作AB.

类似地，称$\bigcap_{k=1}^{n}A_k$为n个事件$A_1,A_2,\dots,A_n$的积事件；称$\bigcap_{k=1}^{\infty}A_k$为可列个事件$A_1,A_2,\dots$的积事件。

差事件：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 发生且 B 不发生时事件$A-B$发生.

考点 6：随机事件的运算律

定义：

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：$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$; $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$.

德摩根律：$\overline{A\cup B}=\bar{A}\cap\bar{B}$; $\overline{A\cap B}=\bar{A}\cup\bar{B}$.

考点 7：频率

定义：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n_A/n称为事件 A 发生的频率,并记成$f_n(A)$.

频率具有下述基本性质:

1．$0\le f_n(A)\le 1$.

2．$f_n(S)=1$.

3．若$A_1,A_2,\dots,A_k$是两两互不相容的事件,则

$$f_n(A_1\cup A_2\cup\dots\cup A_k)=f_n(A_1)+f_n(A_2)+\cdots+f_n(A_k).$$

考点 8：概率

定义： 设 E 是随机试验，S 是它的样本空间. 对于 E 的每一事件 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(·)满足下列条件:

1．非负性：对于每一个事件 A，有 P(A)≥0.

2．规范性：对于必然事件 S，有 P(S)=1.

3．可列可加性：设$A_1,A_2,\dots$是两两互不相容的事件，即对于$A_iA_j=\emptyset,i\neq j,i,j=1,2,\dots$有

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

考点 9：概率的性质

性质 1：概率规范性： $P(\emptyset)=0.$

性质 2：有限可加性： 若$A_1,A_2,\dots,A_n$是两两互不相容的事件,则

$$P(A_1\cup A_2\cup\dots\cup A_n)=P(A_1)+P(A_2)+\cdots+P(A_n).$$

性质 3：概率不等式： 设 A,B 是两事件，若$A\subset B$,则有

$$P(B-A)=P(B)-P(A),$$

$$P(B)\geq P(A).$$

对于任一事件 A，

$$P(A)\leq 1.$$

性质 4：逆事件概率： 对于任一事件 A，

$$P(\bar{A})=1-P(A).$$

性质 5：加法公式： 对于任意两事件 A,B，有

$$P(A\cup B)=P(A)+P(B)-P(AB).$$

性质 6：减法公式： 对于任意两事件 A,B，有

$$P(A\bar{B})=P(A-B)=P(A)-P(AB).$$

补充： 加法公式推广到多个事件

对于任意三个事件 A,B,C,则有

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC).$$

考点 10：等可能模型（古典模型）

定义：对于具有两个共同的特点的试验E_1, E_2:

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.

具有以上两个特点的试验是大量存在的.这种试验称为等可能概型.它在概率论发展初期曾是主要的研究对象,所以也称为古典概型.

考点 11：等可能模型中事件概率

定义：设试验的样本空间为$S = \{e_1, e_2, \dots, e_n\}$. 由于在试验中每个基本事件发生的可能性相同，即有

$$P(\{e_1\}) = P(\{e_2\}) = \cdots = P(\{e_n\}),$$

又由于基本事件是两两互不相容的,于是

$$1 = P(S) = P(\{e_1\} \cup \{e_2\} \dots \cup \{e_n\}) = P(\{e_1\}) + P(\{e_2\}) + \cdots + P(\{e_n\}) = nP(\{e_i\})$$

$$P(\{e_i\}) = \frac{1}{n}, \quad i = 1,2,\dots,n.$$

若事件A包含k个基本事件,即$A = \{e_{i_1}\} \cup \{e_{i_2}\} \dots \cup \{e_{i_k}\}$,这里$i_1, i_2, \dots, i_k$是$1,2,\dots,n$中某$k$个不同的数,则有

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

上式就是等可能概型中事件A的概率的计算公式.

考点 12：条件概率

定义：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.

补充：条件概率$P(\cdot|A)$符合概率定义中的三个条件,即

1．非负性:对于每一事件 B,有$P(B|A) \ge 0$.

2．规范性:对于必然事件 S,有$P(S|A) = 1$.

3．可列可加性:设$B_1, B_2, \dots$是两两互不相容的事件,则有

$$P(\cup_{i=1}^{\infty} B_i \,|A) = \sum_{i=1}^{\infty} P(B_i|A).$$

考点 13：乘法公式

定义：设$P(A)>0$，则有

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

上式称为乘法公式.

补充 1：将上式推广到多个事件的积事件，设 A,B,C 为事件，且 P(AB)>0，则有

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

补充 2：设$A_1,A_2,\cdots,A_n$为多个事件，$n\geq 2$，且$P(A_1,A_2,\cdots,A_{n-1})>0$，则有

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

考点 14：样本空间的划分

定义：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

(i) $B_iB_j=\emptyset, i\neq j, i,j=1,2,\dots,n$

(ii) $B_1\cup B_2\cup\dots\cup B_n=S$,

则称 $B_1,B_2,\dots,B_n$ 为样本空间 S 的一个划分.

补充：若$B_1,B_2,\dots,B_n$是样本空间的一个划分，那么，对每次试验，事件$B_1,B_2,\dots,B_n$中必有一个且仅有一个发生.

考点 15：全概率公式

定义：设试验 E 的样本空间为 S,A 为 E 的事件，$B_1,B_2,\dots,B_n$为 S 的一个划分,且$P(B_i)>0(i=1,2,\dots,n)$,则

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

称为全概率公式.

补充：在很多实际问题中 P(A)不易直接求得,但却容易找到 S 的一个划分$B_1,B_2,\dots,B_n$，且 $P(B_i)$和 $P(A|B_i)$或为已知,或容易求得,那么就可以全概率公式求出 P(A).

考点 16：贝叶斯公式

定义：设试验 E 的样本空间为 S.A 为 E 的事件，$B_1, B_2, \dots, B_n$为 S 的一个划分,且$P(A) > 0, P(B_i) > 0 (i = 1,2, \dots, n)$,则

$$P(B_i|A)=\frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n}P(A|B_j)P(B_j)}, i=1,2,\dots,n.$$

称为贝叶斯公式.

补充：当$n = 2$时，将B_1记为B，此时B_2就是$\bar{B}$，那么全概率公式和贝叶斯公式分别成为

$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$$

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

，，

考点 17：独立性

定义：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,简称 A,B 独立.

性质 1：若 P(A)>0,P(B)>0,则 A,B 相互独立与 A,B 互不相容不能同时成立.

性质 2：设 A,B 是两事件,且$P(A) > 0$.若 A,B 相互独立,则$P(B|A) = P(B)$.反之亦然.

性质 3：若事件 A 与 B 相互独立,则下列各对事件也相互独立:

$$\mathrm{A}与\bar{\mathrm{B}}，\bar{\mathrm{A}}与\mathrm{B}，\bar{\mathrm{A}}与\bar{\mathrm{B}}$$

补充：设 A,B,C 是三个事件，如果满足等式

$$\left.\begin{aligned}P(AB)&=P(A)P(B),\\P(BC)&=P(B)P(C),\\P(AC)&=P(A)P(C),\\P(ABC)&=P(A)P(B)P(C),\end{aligned}\right\}$$

则称事件 A,B,C 相互独立。

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

第二章 随机变量及其分布

考点 18：随机变量

定义：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.

一般以大写的字母如$X,Y,Z,W,\ldots$表示随机变量,而以小写字母工$x,y,z,w,\ldots$表示实数.

补充：随机变量的取值随试验的结果而定,而试验的各个结果出现有一定的概率，因而随机变量的取值有一定的概率.

考点 19：离散型随机变量

定义：随机变量全部可能取到的值是有限个或可列无限多个,这种随机变量称为离散型随机变量.

补充：设离散型随机变量X所有可能取的值为$x_k(k=1,2,\ldots)$,X取各个可能值的概率,即事件$\{X=x_k\}$的概率,为

$$P\{X=x_k\}=p_k,k=1,2,\ldots$$

由概率的定义,p_k满足如下两个条件:

1．$p_k\geq 0,k=1,2,\ldots$

2．$\sum_{k=1}^{\infty}p_k=1$.

考点 20：0-1 分布

定义：设随机变量 X 只可能取 0 与 1 两个值，它的分布律是

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

则称 X 服从以 p 为参数的 $(0-1)$ 分布或两点分布.

$(0-1)$ 分布的分布律也可写成

X	0	1
p_k	$1-p$	p

补充：对于一个随机试验，如果它的样本空间只包含两个元素，即 $S=\{e_1,e_2\}$，我们总能在 S 上定义一个服从 $(0-1)$ 分布的随机变量

$$X=X(e)=\begin{cases}0,当e=e_1,\\1,当e=e_2\end{cases}$$

来描述这个随机试验的结果.

考点 21：伯努利试验、二项分布

定义：设试验E只有两个可能结果：A 及 $\bar{A}$，则称 E 为伯努利试验.

设 $P(A)=p(0<p<1)$，此时 $P(\bar{A})=1-p$. 将 E 独立重复地进行 n 次，则称这一串重复的独立试验为 n 重伯努利试验.

这里“重复”是指在每次试验中 $P(A)=p$ 保持不变;“独立”是指各次试验的结果互不影响.

补充：以 X 表示 n次试验中事件 A 发生的次数，X 所有可能取的值为 $0,1,2,\cdots,n$.若

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

称随机变量 X 服从参数为 n,p 的二项分布，并记为 $X\sim b(n,p)$.

特别，当 $n=1$ 时二项分布化为

$$P\{X=k\}=p^k q^{1-k},\ k=0,1.$$

这就是 $(0-1)$ 分布.

考点 22：泊松分布

定义：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

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

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

易知，$P\{X=k\}\geqslant 0,k=0,1,2,\cdots$，且有

$$\sum_{k=0}^{\infty}P\{X=k\}=\sum_{k=0}^{\infty}\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}=\mathrm{e}^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=\mathrm{e}^{-\lambda}\cdot\mathrm{e}^{\lambda}=1.$$

即 $P\{X=k\}$ 满足条件非负性和正则性.

考点 23：泊松分布的二项近似

定义：(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

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

泊松定理表明当 n 很大，p 很小 $(np=\lambda)$ 时有以下近似式

$$\binom{n}{k}p^k(1-p)^{n-k}\approx\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}\ (\text{其中}\lambda=np).$$

也就是说以 n,p 为参数的二项分布的概率值可以由参数为 $\lambda=np$ 的泊松分布的概率值近似.

考点 24：超几何分布

定义： 设有 N 件产品，其中有 M 件不合格品. 若从中不放回地随机抽取 n 件. 则其中含有的不合格品的件数 X 服从超几何分布，记为 $X \sim h(n, N, M)$. 超几何分布的概率分布列为（见第一章中例 1.2.3）

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, k=0,1,\cdots,r.$$

其中 $r=\min\{M,n\}$，且 $M \leqslant N, n \leqslant N, n, N, M$ 均为正整数.

补充： 超几何分布的二项近似

当 $n \ll N$ 时,即抽取个数 n 远小于产品总数 N 时,每次抽取后，总体中的不合格品率 $p=M/N$ 改变甚微，所以不放回抽样可近似地看成放回抽样，这时超几何分布可用二项分布近似:

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \cong \binom{n}{k} p^k(1-p)^{n-k}, 其中 p=\frac{M}{N}.$$

考点 25：几何分布

定义： 在伯努利试验序列中，记每次试验中事件 A 发生的概率为 p，如果 X 为事件 A 首次出现时的试验次数，则 X 的可能取值为 $1,2,\cdots$，称 X 服从几何分布，记为 $X \sim \mathrm{Ge}(p)$，其分布列为

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

补充： 几何分布的无记忆性

设 $X \sim Ge(p)$，则对任意正整数 m 与 n 有

$$P(X>m+n \mid X>m)=P(X>n).$$

这表明：在前 m 次试验中 A 没有出现的条件下,则在接下去的 n 次试验中 A仍未出现的概率只与 n 有关，而与以前的 m 次试验无关.

考点 26：随机变量的分布函数

定义： 设 X 是一个随机变量，x 是任意实数，函数

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

称为 X 的分布函数.

对于任意实数 $x_1, x_2 (x_1 < x_2)$，有

$$P\{x_1 < X \leqslant x_2\} = P\{X \leqslant x_2\} - P\{X \leqslant x_1\} = F(x_2) - F(x_1),$$

因此，若已知 X 的分布函数，就知道 X 落在任一区间 $(x_1, x_2\ ]$上的概率.

考点 27：分布函数的性质

1.非负性.对于任意实数 $x_1, x_2 (x_1 < x_2)$，有

$$F(x_2) - F(x_1) = P\{x_1 < X \leqslant x_2\} \geqslant 0.$$

2.有界性.$0 \leqslant F(x) \leqslant 1$，且

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0,\ F(\infty) = \lim_{x \to \infty} F(x) = 1.$$

3.右连续.$F(x+0) = F(x)$.

补充： 满足上述三条件，必是某随机变量的分布函数.

考点 28：连续性随机变量

定义： 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

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

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

补充： 连续型随机变量的分布函数是连续函数.

考点 29：概率密度性质

定义：

1.$f(x) \geqslant 0$.

2.$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$.

3.对于任意实数 $x_1, x_2 (x_1 \leqslant x_2)$,

$$P\{x_1 < X \leqslant x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\mathrm{d}x.$$

4.若 $f(x)$ 在点 x 处连续，则有 $\mathrm{F}'(x) = f(x)$.

补充：若 $f(x)$ 具备性质 1、2，引入$G(x)=\int_{-\infty}^{x}f(t)\mathrm{d}t$，它是某一随机变量 X 的分布函数，$f(x)$ 是 X 的概率密度.

连续性随机变量取任一指定实数值 a 的概率均为 0，即 $P\{X=a\}=0$. 事件 $\{X=a\}$ 并非不可能事件，但有 $P\{X=a\}=0$. 这就是说，若 A 是不可能事件，则有 $P(A)=0$；反之，若 $P(A)=0$,并不一定意味着 A 是不可能事件.

随机变量 X 的"概率分布"时，指的是它的分布函数;或者，当 X 是连续型随机变量时，指的是它的概率密度，当 X 是离散型随机变量时,指的是它的分布律.

考点 30：均匀分布

定义：若连续型随机变量 X 具有概率密度

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

则称 X 在区间 (a,b) 上服从均匀分布，记为 $X\sim U(a,b)$.

补充：在区间 (a,b) 上服从均匀分布的随机变量 X，落在区间 (a,b) 中任意等长度的子区间内的可能性是相同的. 或者说落在(a,b) 的子区间内的概率只依赖于子区间的长度而与子区间的位置无关.

考点 31：指数分布

定义：若连续型随机变量 X 的概率密度为

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

其中 $\theta>0$ 为常数,则称 X 服从参数为 θ 的指数分布.

随机变量X 的分布函数为

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

服从指数分布的随机变量 X 具有以下有趣的性质:

补充：指数分布的无记忆性

对于任意 $s,t>0$，有

$$P\{X>s+t\mid X>s\}=P\{X>t\}.$$

如果 X 是某一元件的寿命,X 服从指数分布，则在元件已使用了 s h的条件下，它总共能使用至少 $(s+t)$h 的条件概率，与从开始使用时算起它至少能使用 t h 的概率相等.

考点 32：正态分布

定义： 若连续型随机变量 X 的概率密度为

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

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 $X\sim N(\mu,\sigma^2)$.

补充：

1.曲线关于 $x=\mu$ 对称．这表明对于任意 $h>0$

$$P\{\mu-h<X\leqslant\mu\}=P\{\mu<X\leqslant\mu+h\}.$$

2.当 $x=\mu$ 时取到最大值，此时

$$f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}.$$

3.在 $x=\mu\pm\sigma$ 处曲线有拐点．曲线以 Ox 轴为渐近线.

4.如果固定 σ，改变 μ 的值，则图形沿着 Ox 轴平移，而不改变其形状，可见正态分布的概率密度曲线 $y=f(x)$ 的位置完全由参数 μ 所确定．μ称为位置参数.

如果固定 μ，改变 σ，由于最大值 $f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}$，可知当 σ 越小时图形变得越尖。

考点 33：标准正态分布

定义： 正态分布中，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，即有

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

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

补充 1： $\Phi(-x)=1-\Phi(x)$.

补充 2： 若随机变量 $X\sim N(\mu,\sigma^2)$，则 $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$.

若随机变量 $X\sim N(\mu,\sigma^2)$，则它的分布函数 $F(x)$ 可写成

$$F(x)=P\{X\leqslant x\}=P\left\{\frac{X-\mu}{\sigma}\leqslant\frac{x-\mu}{\sigma}\right\}=\Phi\left(\frac{x-\mu}{\sigma}\right).$$

对于任意区间 $(x_1,x_2]$，有

$$P\{x_1 < X \leqslant x_2\} = P\left\{\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leqslant \frac{x_2-\mu}{\sigma}\right\}$$
$$= \Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right).$$

考点 34：上 α 分位数

定义：设 $X \sim N(0,1)$，若 z_α 满足条件

$$P\{X > z_\alpha\} = \alpha, 0 < \alpha < 1,$$

则称 z_α 为标准正态分布的上 α 分位数.

补充：下面列出了几个常用的 z_α 的值:

α	0.001	0.005	0.01	0.025	0.05	0.10
z_σ	3.090	2.576	2.326	1.960	1.645	1.282

由 $\varphi(x)$ 图形的对称性知道 $z_{1-\alpha} = -z_\alpha$.

考点 35：随机变量的函数的分布——$g(\cdot)$ 是严格单调函数的情况

定义：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.

第三章 多维随机变量及其分布

考点 36：二维随机变量

定义： 设 E 是一个随机试验，它的样本空间是 $S=\{e\}$，设 $X=X(e)$ 和 $Y=Y(e)$ 是定义在 S 上的随机变量，由它们构成的一个向量 (X,Y),叫做二维随机向量或二维随机变量.

考点 37：二维随机变量的分布函数

定义： 设 (X,Y) 是二维随机变量，对于任意实数 x,y，二元函数:

$$F(x,y)=P\{(X\leqslant x)\cap(Y\leqslant y)\}\overset{\text{记成}}{=}PX\leqslant x,Y\leqslant y\}$$

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.

考点 38：二维随机变量的分布函数的基本性质

定义：

1.单调性

$F(x,y)$ 是变量 x 和 y 的不减函数，即对于任意固定的 y，当 $x_2>x_1$ 时 $F(x_2,y)\geqslant F(x_1,y)$; 对于任意固定的 x，当 $y_2>y_1$ 时 $F(x,y_2)\geqslant F(x,y_1)$.

2.有界性

$0\leqslant F(x,y)\leqslant 1$，且

对于任意固定的 $y,F(-\infty,y)=0$,

对于任意固定的 $x,F(x,-\infty)=0$,

$F(-\infty,-\infty)=0,F(\infty,\infty)=1$.

3.右连续性

$F(x+0,y)=F(x,y),F(x,y+0)=F(x,y)$，即 $F(x,y)$ 关于 x 右连续,关于 y 也右连续.

4.非负性

对于任意 $(x_1,y_1),(x_2,y_2),x_1<x_2,y_1<y_2$，下述不等式成立:

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

考点 39：二维离散型随机变量

定义： 如果二维随机变量 (X,Y) 全部可能取到的值是有限对或可列无限多对，则称 (X,Y) 是二维离散型随机变量.

补充： 设二维离散型随机变量 (X,Y) 所有可能取的值为 $(x_i,y_j),i,j=1,2,\cdots$，记 $P\{X=x_i,Y=y_j\}=p_{ij},i,j=1,2,\cdots$，则由概率的定义有

$$p_{ij}\geqslant 0,\ \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}p_{ij}=1.$$

称 $P\{X=x_i,Y=y_j\}=p_{ij},i,j=1,2,\cdots$ 为二维离散型随机变量 (X,Y) 的分布律，或称为随机变量 X 和 Y 的联合分布律.

用表格来表示 X 和 Y 的联合分布律.

Y X	x_1	x_2	$\cdots$	x_i	$\cdots$
y_1	p_{11}	p_{21}	$\cdots$	p_{i1}	$\cdots$
y_2	p_{12}	p_{22}	$\cdots$	p_{i2}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	
y_j	p_{1j}	p_{2j}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	

考点 40：二维连续型随机变量

定义： 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

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

则称 (X,Y) 是**二维连续型随机变量**，函数 $f(x,y)$ 称为二维连续型随机变量 (X,Y) 的概率密度，或称为随机变量 X 和 Y 的**联合概率密度**.

补充： 概率密度 $f(x,y)$ 具有以下性质：

1.$f(x,y)\geqslant 0$.

2.$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\mathrm{d}x\,\mathrm{d}y=F(\infty,\infty)=1$.

3.设 G 是 xOy 平面上的区域，点 (X,Y) 落在 G 内的概率为

$$P\{(X,Y)\in G\}=\iint_G f(x,y)\mathrm{d}x\,\mathrm{d}y.$$

4. 若 $f(x,y)$ 在点 (x,y) 连续，则有

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

考点 41：二维均匀分布

定义：设 D 为 $\mathbf{R}^n$ 中的一个有界区域，其度量(平面的为面积，空间的为体积等)为 S_D，如果多维随机变量 $(X_1,X_2,\cdots,X_n)$ 的联合密度函数为

$$p(x_1,x_2,\cdots,x_n)=\begin{cases}\dfrac{1}{S_D}, & (x_1,x_2,\cdots,x_n)\in D,\\ 0, & \text{其他}.\end{cases}$$

则称 $(X_1,X_2,\cdots,X_n)$ 服从 D 上的多维均匀分布，记为 $(X_1,X_2,\cdots,X_n)\sim U(D)$.

补充：二维均匀分布所描述的随机现象就是向平面区域 D 中随机投点,如果该点坐标 (X,Y) 落在 D 的子区域 G 中的概率只与 G 的面积有关，而与 G 的位置无关，即为几何概率．现由二维均匀分布来描述，则

$$P((X,Y)\in G)=\iiint_G p(x,y)\mathrm{d}x\,\mathrm{d}y=\iint_G \frac{1}{S_D}\,\mathrm{d}x\,\mathrm{d}y=\frac{G\text{的面积}}{D\text{的面积}}.$$

考点 42：二维正态分布

定义：若二维随机变量 (X,Y) 的联合密度函数为

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}\right.\right.$$
$$\left.\left.-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\},-\infty<x,y<\infty,$$

则称 (X,Y) 服从二元正态分布，记为 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$．其中五个参数的取值范围分别是

$$-\infty<\mu_1,\mu_2<\infty,\ \sigma_1,\sigma_2>0,\ -1\leqslant\rho\leqslant 1.$$

其中 μ_1,μ_2 分别是 X 与 Y 的均值，σ_1^2,σ_2^2 分别是 X 与 Y 的方差，ρ 是 X 与 Y 的相关系数.

考点 43：n 维随机变量及其分布

定义：设 E 是一个随机试验，它的样本空间是 $S=\{e\}$，设 $X_1=X_1(e),X_2=X_2(e),\cdots,X_n=X_n(e)$ 是定义在 S 上的随机变量，由它们构成的一个 n 维向量 $(X_1,X_2,\cdots,X_n)$ 称为 n 维随机向量或 $\boldsymbol{n}$ **维随机变量**.

补充：对于任意 n 个实数 $x_1,x_2,\cdots,x_n,n$ 元函数

$$F(x_1,x_2,\cdots,x_n)=P\{X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_n\leqslant x_n\}$$

称为 n 维随机变量 $(X_1,X_2,\cdots,X_n)$ 的分布函数或随机变量 $X_1,X_2,\cdots,X_n$ 的联合分布函数. 它具有类似于二维随机变量的分布函数的性质.

考点 44：边缘分布函数

定义：二维随机变量 (X,Y)分布函数记为 $F(x,y)$. 随机变量 X 和 Y 各自的分布函数分别记为 $F_X(x),F_Y(y)$，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. 边缘分布函数可以由 (X,Y) 的分布函数 $F(x,y)$ 所确定，

$$F_X(x)=P\{X\leqslant x\}=P\{X\leqslant x,Y<\infty\}=F(x,\infty),$$

即

$$F_X(x)=F(x,\infty).$$

就是说，只要在函数 $F(x,y)$ 中令 $y\to\infty$ 就能得到 $F_X(x)$.

同理

$$F_Y(y)=F(\infty,y).$$

考点 45：二维离散型随机变量的边缘分布函数

定义：对于离散型随机变量 (X,Y)， X 的分布律为

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

同样，Y 的分布律为

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

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

考点 46：二维连续型随机变量的边缘分布函数

定义： 对于连续型随机变量 (X,Y)，设它的概率密度为 $f(x,y)$，由于

$$F_X(x)=F(x,\infty)=\int_{-\infty}^{x}\left[\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y\right]\mathrm{d}x,$$

X 是一个连续型随机变量，且其概率密度为

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

同样，Y 也是一个连续型随机变量，其概率密度为

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

分别称 $f_X(x),f_Y(y)$ 为 (X,Y) 关于 X 和关于 Y 的**边缘概率密度**.

考点 47：二维离散型随机变量的条件分布

定义： 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

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

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

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

为在 $X=x_i$ 条件下随机变量 Y 的**条件分布律.**

补充：

1.$\mathrm{P}\{X=x_i\mid Y=y_j\}\geqslant 0.$

2. $\sum_{i=1}^{\infty}P\{X=x_i\mid Y=y_j\}=\sum_{i=1}^{\infty}\frac{p_{ij}}{p_{\cdot j}}=\frac{1}{p_{\cdot j}}\sum_{i=1}^{\infty}p_{ij}=\frac{p_{\cdot j}}{p_{\cdot j}}=1.$

考点 48：二维连续型随机变量的条件分布

定义： 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

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

称

$$\int_{-\infty}^{x}f_{X|Y}(x\mid y)\mathrm{d}x=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}\mathrm{d}x$$

为在 $Y=y$ 的条件下 X 的条件分布函数,记为 $P\{X\leqslant x\mid Y=y\}$ 或 $F_{X|Y}(x\mid y)$.

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

考点 49：随机变量之间的独立性

定义： 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. 若对于所有 x,y 有

$$P\{X\leqslant x,Y\leqslant y\}=P\{X\leqslant x\}P\{Y\leqslant y\},$$

即

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

则称随机变量 X 和 Y 是相互独立的.

补充 1： 设 (X,Y) 是连续型随机变量，$f(x,y),f_X(x),f_Y(y)$ 分别为 (X,Y) 的概率密度和边缘概率密度，则 X 和 Y 相互独立的条件等价于:

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

补充 2： 当 (X,Y) 是离散型随机变量时，X 和 Y 相互独立的条件等价于: 对于 (X,Y) 的所有可能取的值 $(x_i，y_j)$ 有

$$P\{X=x_i，Y=y_j\}=P\{X=x_i\}P\{Y=y_j\}.$$

补充 3： 对于二维正态随机变量(X,Y),X和Y相互独立的充要条件是参数$\boldsymbol{\rho}=\mathbf{0}$(ρ为相关系数).

补充 4： 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.

考点 50：两个随机变量的函数的分布——Z = X + Y 的分布

定义： 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

$$f_{X+Y}(z)=\int_{-\infty}^{\infty}f(z-y,y)\mathrm{d}y,$$

或

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

又若 X 和 Y 相互独立，设 (X,Y) 关于 X,Y 的边缘概率密度分别为 $f_X(x)$，$f_Y(y)$，则

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

和

$$f_{X+Y}(z)=\int_{-\infty}^{\infty}f_X(x)f_Y(z-x)\mathrm{d}x.$$

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

$$f_X * f_Y = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) \mathrm{d}y = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) \mathrm{d}x.$$

补充：设 X,Y 相互独立且 $X \sim N(\mu_1,\sigma_1^2), Y \sim N(\mu_2,\sigma_2^2)$. 则 $Z = X+Y$ 仍然服从正态分布,且有 $Z \sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

n 个独立正态随机变量之和的情况. 即若 $X_i \sim N\left(\mu_i,\sigma_i^2\right)(i=1,2,\cdots,n)$, 且它们相互独立, 则它们的和 $Z = X_1+X_2+\cdots+X_n$ 仍然服从正态分布, 且有 $Z \sim N(\mu_1+\mu_2+\cdots+\mu_n,\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2)$.

更一般地,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.

考点 51：两个随机变量的函数的分布——$Z=\frac{Y}{X}$ 的分布、 $Z=XY$ 的分布

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

$$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f(x,xz) \mathrm{d}x,$$
$$f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f\left(x,\frac{z}{x}\right) \mathrm{d}x.$$

又若 X 和 Y 相互独立. 设 (X,Y) 关于 X,Y 的边缘概率密度分别为 $f_X(x)$, $f_Y(y)$, 则

$$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz) \mathrm{d}x.$$
$$f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y\left(\frac{z}{x}\right) \mathrm{d}x.$$

考点 52：两个随机变量的函数的分布——$M=\max\{X,Y\}$ 及 $N=\min\{X,Y\}$ 的分布

定义：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$F_Y(y)$.

由于 X 和 Y 相互独立, 得到 $M=\max\{X,Y\}$ 的分布函数为

$$F_{\max}(z) = P\{M \leqslant z\} = P\{X \leqslant z, Y \leqslant z\} = P\{X \leqslant z\} P\{Y \leqslant z\} = F_X(z) F_Y(z)..$$

$N=\min\{X,Y\}$ 的分布函数为

$$\begin{aligned} F_{\min}(z) &= P\{N \leqslant z\} = 1 - P\{N > z\} \\ &= 1 - P\{X > z, Y > z\} = 1 - P\{X > z\} P\{Y > z\}. \\ &= 1 - [1 - F_X(z)][1 - F_Y(z)]. \end{aligned}$$

补充：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

$$F_{\max}(z) = F_{X_1}(z) F_{X_2}(z) \cdots F_{X_n}(z),$$
$$F_{\min}(z) = 1 - \left[1 - F_{X_1}(z)\right]\left[1 - F_{X_2}(z)\right] \cdots \left[1 - F_{X_n}(z)\right].$$

特别，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

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

第四章 随机变量的数字特征

考点 53：随机变量的数学期望

定义：设离散型随机变量 X 的分布律为

$$P\{X=x_k\}=p_k,\ k=1,2,\cdots.$$

若级数

$$\sum_{k=1}^{\infty}x_kp_k$$

绝对收敛，则称级数 $\sum_{k=1}^{\infty}x_kp_k$ 的和为随机变量 X 的数学期望,记为 $E(X)$. 即

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

定义：设连续型随机变量 X 的概率密度为 $f(x)$，若积分

$$\int_{-\infty}^{\infty}xf(x)\mathrm{d}x$$

绝对收敛，则称积分 $\int_{-\infty}^{\infty}xf(x)\mathrm{d}x$ 的值为随机变量 X 的数学期里，记为 $E(X)$. 即

$$E(X)=\int_{-\infty}^{\infty}xf(x)\mathrm{d}x.$$

数学期望简称期望，又称为均值.

考点 54：随机变量的函数的数学期望

定义：设 Y 是随机变量 X 的函数: $Y=g(X)$（ g 是连续函数).

1.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 $\sum_{k=1}^{\infty}g(x_k)p_k$ 绝对收敛，则有

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

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

$$E(Y)=E[g(X)]=\int_{-\infty}^{\infty}g(x)f(x)\mathrm{d}x$$

补充：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

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

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.

考点 55：随机变量数学期望的性质

定义：

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变量之积的情况.

考点 56：随机变量的方差

定义：设 X 是随机变量，若 $E\{[X-E(X)]^2\}$ 存在，则称它为 X 的方差，记为 $D(X)$ 或 $\mathrm{Var}(X)$，即

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

将$\sqrt{D(X)}$记为 $\sigma(X)$，称为标准差或均方差.

对于离散型随机变量，

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

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

其中 $f(x)$ 是 X 的概率密度.

补充：随机变量 X 的方差可按下列公式计算:

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

考点 57：随机变量方差的性质

定义：

1.设 C 是常数，则 $D(C)=0$.

2.设 X 是随机变量，C 是常数，则有

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

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，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

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

补充 1：若 $X_i\sim N(\mu_i,\sigma_i^2),i=1,2,\cdots,n$，且它们相互独立，则它们的线性组合：$C_1X_1+C_2X_2+\cdots+C_nX_n(C_1,C_2,\cdots,C_n$ 是不全为 0 的常数)仍然服从正态分布,于是由数学期望和方差的性质可知.

$$C_1X_1+C_2X_2+\cdots+C_nX_n\sim N\left(\sum_{i=1}^{n}C_i\mu_i,\sum_{i=1}^{n}C_i^2\sigma_i^2\right).$$

补充 2：0-1 分布的期望与方差

设随机变量 X 具有 $(0-1)$ 分布，其分布律为

$$P\{X=0\}=1-p,\ P\{X=1\}=p.$$

则

$$E(X)=p$$

$$D(X)=p(1-p).$$

补充 3：二项分布的期望与方差

设随机变量 X 服从$B(n,p)$.

则

$$E(X)=\mathrm{n}p$$

$$D(X)=\mathrm{n}p(1-p).$$

补充 4：泊松分布的期望与方差

设随机变量 $X\sim\pi(\lambda)$

则

$$E(X)=\lambda$$

$$D(X)=\lambda.$$

补充 5：均匀分布的期望与方差

设随机变量 $X\sim U(a,b)$

则

$$E(X)=\frac{a+b}{2}$$

$$D(X)=\frac{(b-a)^2}{12}$$

补充 6：指数分布的期望与方差

设随机变量 X 服从指数分布，其概率密度为

$$f(x)=\begin{cases}\dfrac{1}{\theta}\mathrm{e}^{-x/\theta}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$$

其中 $\theta>0$,则

$$E(X)=\theta$$

$$D(X)=\theta^2.$$

补充 7：正态分布的期望与方差

随机变量 $X\sim N(\mu,\sigma^2)$

$$E(X)=\mu$$
$$D(X)=\sigma^2$$

考点 58：随机变量的协方差

定义：将 $E\{[X-E(X)][Y-E(Y)]\}$ 称为随机变量 X 与 Y 的协方差. 记为 $\mathrm{Cov}\,(X,Y)$，即

而

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

补充：将 $\mathrm{Cov}\,(X,Y)$ 的定义式展开，易得

$$\mathrm{Cov}\,(X,Y)=E(XY)-E(X)E(Y).$$

考点 59：随机变量协方差的性质:

定义：

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

2.$\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$.

3.$\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$, $\mathrm{Cov}(X,X)=D(X)$.

4.$D(X+Y)=D(X)+D(Y)+2\mathrm{Cov}(X,Y)$.

考点 60：随机变量的相关系数

定义：将

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

称为随机变量 X 与 Y 的相关系数.

补充：反算公式

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

考点 61：随机变量相关系数的性质

定义：

1.$|\rho_{XY}|\leqslant 1$.

2.$|\rho_{XY}|=1$ 的充要条件是，存在常数 a,b 使

$$P\{Y=a+bX\}=1.$$

补充 1： 当 $|\rho_{XY}|$ 较大时，我们通常说 X,Y 线性相关的程度较好；当 $|\rho_{XY}|$ 较小时，我们说，X,Y 线性相关的程度较差.特别当 $|\rho_{XY}|=1$ 时,X,Y 之间以概率 1 存在着线性关系.

当 $\rho_{XY}=0$ 时,称 X 和 Y 不相关.

补充 2：假设随机变量 X,Y 的相关系数 ρ_{XY} 存在. 当 X 和 Y 相互独立时， $\mathrm{Cov}(X,Y)=0$，从而 $\rho_{XY}=0$，即 X,Y 不相关. 反之，若 X,Y 不相关，X 和 Y 却不一定相互独立

补充 3：当 (X,Y) 服从二维正态分布时，X 和 Y 不相关与 X和 Y 相互独立是等价的.

补充 4：二维正态随机变量 (X,Y) 的概率密度中的参数 ρ 就是 X 和 Y 的相关系数.

考点 62：随机变量的矩

定义：设 X 和 Y 是随机变量，若

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

存在，称它为 X 的 k 阶原点矩，简称 k 阶矩.

定义：若

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

存在，称它为 X 的 k 阶中心矩.

定义：若

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

存在，称它为 X 和 Y 的 $k+l$ 阶混合矩.

定义：若

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

存在，称它为 X 和 Y 的 $k+l$ 阶混合中心矩.

补充：X 的数学期望 $E(X)$ 是 X 的一阶原点矩，方差 $D(X)$ 是 X 的二阶中心矩，协方差 $\mathrm{Cov}\,(X,Y)$ 是 X 和 Y 的二阶混合中心矩.

考点 63：随机变量的协方差矩阵

定义：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

$$\begin{aligned}c_{11}&=E\{[X_1-E(X_1)]^2\},\\ c_{12}&=E\{[X_1-E(X_1)][X_2-E(X_2)]\},\\ c_{21}&=E\{[X_2-E(X_2)][X_1-E(X_1)]\},\\ c_{22}&=E\{[X_2-E(X_2)]^2\}.\end{aligned}$$

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.

定义：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

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

为 n 维随机变量 $(X_1,X_2,\cdots,X_n)$ 的协方差矩阵.

考点 64：n 维正态随机变量的概率密度

定义：二维正态随机变量 (X_1,X_2) 的概率密度为

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}.$$

引入如下的列矩阵

$$\boldsymbol{X}=\begin{pmatrix}x_1\\x_2\end{pmatrix},\ \boldsymbol{\mu}=\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix}.$$

(X_1,X_2) 的协方差矩阵为

$$\boldsymbol{C}=\begin{pmatrix}c_{11}&c_{12}\\c_{21}&c_{22}\end{pmatrix}=\begin{pmatrix}\sigma_1^2&\rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2&\sigma_2^2\end{pmatrix},$$

它的行列式 $\det\boldsymbol{C}=\sigma_1^2\sigma_2^2(1-\rho^2),\boldsymbol{C}$ 的逆矩阵为

$$\boldsymbol{C}^{-1}=\frac{1}{\det\boldsymbol{C}}\begin{pmatrix}\sigma_2^2&-\rho\sigma_1\sigma_2\\-\rho\sigma_1\sigma_2&\sigma_1^2\end{pmatrix}.$$

于是 (X_1,X_2) 的概率密度可写成

$$f(x_1,x_2)=\frac{1}{(2\pi)^{2/2}(\det\boldsymbol{C})^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right\}.$$

定义： n 维正态随机变量 $(X_1,X_2,\cdots,X_n)$ 的情况.

引入列矩阵

$$\boldsymbol{X}=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\text{和}\boldsymbol{\mu}=\begin{pmatrix}\mu_1\\\mu_2\\\vdots\\\mu_n\end{pmatrix}=\begin{pmatrix}E(X_1)\\E(X_2)\\\vdots\\E(X_n)\end{pmatrix}.$$

n 维正态随机变量 $(X_1,X_2,\cdots,X_n)$ 的概率密度定义为

$$f(x_1,x_2,\cdots,x_n)=\frac{1}{(2\pi)^{n/2}(\det\boldsymbol{C})^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right\},$$

其中 C 是 $(X_1,X_2,\cdots,X_n)$ 的协方差矩阵.

考点 65：n 维正态随机变量的性质

定义：

1.n 维正态随机变量 $(X_1,X_2,\cdots,X_n)$ 的每一个分量 $X_i,i=1,2,\cdots,n$ 都是正态随机变量；反之，若 $X_1,X_2,\cdots,X_n$ 都是正态随机变量，且相互独立，则（X_1，$X_2,\cdots,X_n$）是 n 维正态随机变量.

2.n 维随机变量 $(X_1,X_2,\cdots,X_n)$ 服从 n 维正态分布的充要条件是 X_1，$X_2,\cdots,X_n$ 的任意的线性组合

$$l_1X_1+l_2X_2+\cdots+l_nX_n$$

服从一维正态分布(其中 $l_1, l_2, \cdots, l_n$ 不全为零).

3.若 $(X_1, X_2, \cdots, X_n)$ 服从 n 维正态分布，设 $Y_1, Y_2, \cdots, Y_k$ 是 $X_j(j=1,2,\cdots,n)$ 的线性函数，则 $(Y_1, Y_2, \cdots, Y_k)$ 也服从多维正态分布.

这一性质称为正态变量的线性变换不变性.

4.设 $(X_1, X_2, \cdots, X_n)$ 服从 n 维正态分布，则 "$X_1, X_2, \cdots, X_n$ 相互独立" 与 "$X_1, X_2, \cdots, X_n$ 两两不相关"是等价的.

第五章 大数定理及中心极限

考点 66：依概率收敛

定义： 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 ε，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

$$Y_n \xrightarrow{P} a.$$

补充： 设 $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$，又设函数 $g(x,y)$ 在点 (a,b) 连续，则

$$g(X_n, Y_n) \xrightarrow{P} g(a,b).$$

考点 67：伯努利大数定律

定义： 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 $\varepsilon > 0$，有

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

或

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

补充： 伯努利大数定律的结果表明，对于任意 $\varepsilon > 0$，只要重复独立试验的次数 n充分大，事件 $\left\{\left|\frac{f_A}{n} - p\right| \geqslant \varepsilon\right\}$ 是一个小概率事件，由实际推断原理知，这一事件实际上几乎是不发生的，即在 n 充分大时事件 $\left\{\left|\frac{f_A}{n} - p\right| < \varepsilon\right\}$ 实际上几乎是必定要发生的，亦即对于给定的任意小的正数 ε，在 n 充分大时，事件 "频率 $\frac{f_A}{n}$ 与概率 p 的偏差小于 ε "实际上几乎是必定要发生的. 也即频率稳定性的真正含义. 由实际推断原理，在实际应用中，当试验次数很大时，便可以用事件的频率

来代替事件的概率.

考点 68：切比雪夫大数定理

定义：（切比雪夫大数定律）设 $\{X_n\}$ 为一列两两不相关的随机变量序列，若每个 X_i 的方差存在，且有共同的上界，即 $\mathrm{Var}(X_i) \leqslant c, i = 1,2,\cdots$，则 $|X_n|$ 服从大数定律，即对任意的 $\varepsilon > 0$,

$$\lim_{n\to =} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right) = 1.$$

补充：设随机变量 X 具有数学期望 $E(X) = \mu$，方差 $D(X) = \sigma^2$，则对于任意正数 ε,不等式

$$P\{|X-\mu| \geqslant \varepsilon\} \leqslant \frac{\sigma^2}{\varepsilon^2}$$

成立.这一不等式称为切比雪夫不等式.

切比雪夫不等式也可以写成如下的形式:

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

考点 69：辛钦大数定律

定义：辛钦大数定律 设$X_1, X_2, \ldots$是相互独立,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 个变量的算术平均 $\frac{1}{n}\sum_{k=1}^{n} X_k$，则对于任意 $\varepsilon > 0$，有

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

补充：辛钦大数定律表明，当 $n \to \infty$ 时这个事件的概率趋于 1. 即对于任意正数 ε，当 n 充分大时，不等式 $\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon$成立的概率很大. 通俗地说，辛钦大数定律是说，对于独立同分布且具有均值 μ 的随机变量 $X_1, X_2, \cdots, X_n$，当 n 很大时它们的算术平均 $\frac{1}{n}\sum_{k=1}^{n} X_k$ 很可能接近于 μ.

考点 70：独立同分布的中心极限定理

定义：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

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \leqslant x\right\}$$
$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2}\,\mathrm{d}t = \Phi(x).$$

补充： 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 $\sum_{k=1}^{n} X_k$ 的标准化变量，当 n 充分大时，有

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \overset{\text{近似地}}{\sim} N(0,1).$$

将上式左端改写成 $\frac{\frac{1}{n}\sum_{k=1}^{n} X_k - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$，即当 n充分大时，

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \overset{\text{近似地}}{\sim} N(0,1)\text{或}\bar{X} \overset{\text{近似地}}{\sim} N(\mu, \sigma^2/n).$$

表明均值为 μ,方差为 $\sigma^2 > 0$ 的独立同分布的随机变量 $X_1, X_2, \cdots, X_n$ 的算术平均 $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$,当 n 充分大时近似地服从均值为 μ，方差为 σ^2/n 的正态分布.

考点 71：林德伯格—列维（Lindeberg-Levy）定理

定义：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

$$E(X_k) = \mu_k,\ D(X_k) = \sigma_k^2 > 0,\ k = 1,2,\cdots,$$

记

$$B_n^2 = \sum_{k=1}^{n} \sigma_k^2.$$

若存在正数 δ，使得当 $n \to \infty$ 时，

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

则随机变量之和 $\sum_{k=1}^{n} X_k$ 的标准化变量

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

的分布函数 $F_n(x)$ 对于任意 x，满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n} \leqslant x\right\}$$
$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2}\,\mathrm{d}t = \Phi(x).$$

补充：林德伯格—列维（Lindeberg-Levy）定理表明， 在定理的条件下，随机变量

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

当 n 很大时，近似地服从正态分布 $N(0,1)$. 由此，当 n 很大时，$\sum_{k=1}^{n} X_k = B_n Z_n + \sum_{k=1}^{n} \mu_k$近似地服从正态分布 $N\left(\sum_{k=1}^{n} \mu_k, B_n^2\right)$. 这就是说，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 $\sum_{k=1}^{n} X_k$ 当 n 很大时，就近似地服从正态分布.

考点 72：棣莫弗一拉普拉斯 (De Moivre-Laplace) 定理

定义：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

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

补充：棣莫弗一拉普拉斯定理表明,正态分布是二项分布的极限分布. 当 n 充分大时，可以利用上式来计算二项分布的概率.