概率统计

大数定律和中心极限定理的定义

大数定律（Law of Large Numbers）：

大数定律是一个统计学定理，它描述了**当样本容量增加时，样本平均值会逐渐收敛于总体均值。**具体来说，如果从一个总体中抽取的**独立同分布**的随机样本足够大，那么这些样本的平均值将趋近于总体的均值。大数定律有两个主要形式：弱大数定律和强大数定律。**弱大数定律要求样本的期望存在，而强大数定律要求样本的方差也存在。**

定理 4.3.1（伯努利大数定律） 设 S_n 为 n 重伯努利试验中事件 A 发生的次数，p 为每次试验中 A 出现的概率，则对任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n}-p\right|<\varepsilon\right)=1.$$

那么伯努利大数定律的结论为：对任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)\right|<\varepsilon\right)=1. \tag{4.3.5}$$

定义 4.3.1 设有一随机变量序列 $\{X_n\}$，假如它具有形如(4.3.5)式的性质，则称该随机变量序列 $\{X_n\}$ **服从大数定律**.

定理 4.3.2（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 $\varepsilon>0$，(4.3.5)式成立.

证明 因为 $\{X_n\}$ 两两不相关，故

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i)\leqslant\frac{c}{n}.$$

再由切比雪夫不等式得到：对任意的 $\varepsilon>0$，有

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)\right|<\varepsilon\right)\geqslant 1-\frac{\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)}{\varepsilon^2}\geqslant 1-\frac{c}{n\varepsilon^2}.$$

于是当 $n\to\infty$ 时，有

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

注意，切比雪夫大数定律只要求 $\{X_n\}$ 互不相关，并不要求它们是同分布的.因此，我们很容易推出：如果 $\{X_n\}$ 是独立同分布的随机变量序列，且方差有限，则 $\{X_n\}$ 必定服从大数定律.伯努利大数定律是切比雪夫大数定律的特例.

注意到以上大数定律的证明中，只要有

$$\frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \to 0, \tag{4.3.6}$$

则大数定律就能成立.这个条件(4.3.6)被称为**马尔可夫条件**.

定理 4.3.3(马尔可夫大数定律) 对随机变量序列$\{X_n\}$,若(4.3.6)式成立,则$\{X_n\}$服从大数定律,即对任意的$\varepsilon>0$,(4.3.5)式成立.

证明 利用切比雪夫不等式即可证得.

马尔可夫大数定律的重要性在于:对$\{X_n\}$已经没有任何同分布、独立性、不相关的假定.切比雪夫大数定律显然可由马尔可夫大数定律推出.

我们已经知道,一个随机变量的方差存在,则其数学期望必定存在;但反之不成立,即一个随机变量的数学期望存在,则其方差不一定存在.以上几个大数定律均假设随机变量序列$\{X_n\}$的方差存在,以下的辛钦大数定律去掉了这一假设,仅设每个X_i的数学期望存在,但同时要求$\{X_n\}$为独立同分布的随机变量序列.伯努利大数定律也是辛钦大数定律的特例.

定理 4.3.4(辛钦大数定律) 设$\{X_n\}$为一独立同分布的随机变量序列,若X_i的数学期望存在,则$\{X_n\}$服从大数定律,即对任意的$\varepsilon>0$,(4.3.5)式成立.

辛钦大数定律提供了求随机变量数学期望$E(X)$的近似值的方法:设想对随机变量X独立重复地观察n次,第k次观察值为X_k,则$X_1,X_2,\cdots,X_n$应该是相互独立的,且它们的分布应该与X的分布相同.所以,在$E(X)$存在的条件下,按照辛钦大数定律,当n足够大时,可以把平均观察值

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

作为$E(X)$的近似值.这种做法的一个优点是我们可以不必去管X的分布究竟是怎样的,我们的目的只是寻求数学期望的近似值.

中心极限定理（Central Limit Theorem）：

中心极限定理是一个重要的统计学定理，它描述了**在满足一定条件下，当从任何总体中抽取足够大的样本时，样本均值的分布会趋向于正态分布。**具体来说，中心极限定理有几个不同的形式，其中最著名的是针对**独立同分布**的随机变量的情况。这个定理表明，无论总体的分布是什么，只要样本容量足够大，样本均值的分布都会接近于正态分布。这对于许多统计推断和假设检验问题非常重要，因为正态分布在统计学中有广泛的应用。

定理 4.4.1(林德伯格-莱维(Lindeberg-Lévy)中心极限定理) 设$\{X_n\}$是独立同分布的随机变量序列,且$E(X_i)=\mu,\mathrm{Var}(X_i)=\sigma^2>0$存在,若记

$$Y_n^* = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}},$$

则对任意实数y,有

$$\lim_{n\to\infty} P(Y_n^* \leqslant y) = \Phi(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} \mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t. \tag{4.4.1}$$

定理 4.4.2(棣莫弗-拉普拉斯(de Moivre-Laplace)中心极限定理) 设 n 重伯努利试验中,事件 A 在每次试验中出现的概率为 p $(0<p<1)$,记 S_n 为 n 次试验中事件 A 出现的次数,且记

$$Y_n^* = \frac{S_n - np}{\sqrt{npq}}.$$

则对任意实数 y,有

$$\lim_{n\to\infty} P(Y_n^* \leqslant y) = \Phi(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{t^2}{2}} dt.$$

全概率公式和贝叶斯公式

全概率公式（Law of Total Probability）：

全概率公式用于计算一个事件的概率，通过考虑该事件在不同情境或条件下的发生概率。假设事件 A 可以被划分为多个互不相交的情境或条件（B_1，B_2，B_3，...，B_n），全概率公式表示为：
$P(A) = \Sigma[P(A|B_i) * P(B_i)]$

其中，$P(A)$是事件 A 的概率，$P(A|B_i)$是在条件 B_i 下事件 A 的条件概率，$P(B_i)$是条件 B_i 发生的概率。这个公式允许我们通过考虑不同的情境或条件来计算事件 A 的总体概率。

性质 1.4.3(全概率公式) 设 $B_1, B_2, \cdots, B_n$ 为样本空间 Ω 的一个分割(见图 1.4.2),即 $B_1, B_2, \cdots, B_n$ 互不相容,且 $\bigcup_{i=1}^{n} B_i = \Omega$,如果 $P(B_i) > 0, i = 1, 2, \cdots, n$,则对任一事件 A 有

$$P(A) = \sum_{i=1}^{n} P(B_i) P(A \mid B_i). \tag{1.4.4}$$

证明 因为

$$A = A\Omega = A\left(\bigcup_{i=1}^{n} B_i\right) = \bigcup_{i=1}^{n} (AB_i),$$

且 $AB_1, AB_2, \cdots, AB_n$ 互不相容,所以由可加性得

$$P(A) = P\left(\bigcup_{i=1}^{n} (AB_i)\right) = \sum_{i=1}^{n} P(AB_i),$$

再将 $P(AB_i) = P(B_i)P(A|B_i), i = 1, 2, \cdots, n$,代入上式即得(1.4.4).

贝叶斯公式（Bayes' Theorem）：

贝叶斯公式是用于更新概率分布的重要工具，特别是在统计推断和机器学习中。它用于计算在已知某一条件下另一个事件的概率。贝叶斯公式的一般形式如下：$P(A|B) = [P(B|A) * P(A)]/P(B)$

其中，$P(A|B)$是在条件 B 下事件 A 的概率（后验概率），$P(B|A)$是在条件 A 下事件 B 的概率（似然性），$P(A)$和$P(B)$分别是事件 A 和事件 B 的先验概率，其中$P(B)$通常可以计算为全概率公式中的求和项。

性质 1.4.4（贝叶斯公式） 设 $B_1, B_2, \cdots, B_n$ 是样本空间 Ω 的一个分割，即 $B_1, B_2, \cdots, B_n$ 互不相容，且 $\bigcup_{i=1}^{n} B_i = \Omega$，如果 $P(A) > 0, P(B_i) > 0, i = 1, 2, \cdots, n$，则

$$P(B_i \mid A) = \frac{P(B_i)P(A \mid B_i)}{\sum_{j=1}^{n} P(B_j)P(A \mid B_j)}, \quad i = 1, 2, \cdots, n. \tag{1.4.6}$$

先验概率和后验概率

先验概率（Prior Probability）：

先验概率是指在考虑任何新的数据或信息之前，根据先前的知识、经验或信息所赋予的概率。它反映了我们对某个事件或假设的初始信念或估计，通常以$P(A)$表示，其中 A 是事件或假设。先验概率基于以往的观察或领域知识，可以是主观的或客观的。在贝叶斯统计中，先验概率在贝叶斯推断中起到重要作用，它用于建立初始概率分布，然后根据新的数据来更新概率分布，得到后验概率。

后验概率（Posterior Probability）：

后验概率是在考虑了新的数据或信息之后，根据贝叶斯定理计算得出的概率。它表示在考虑了新的证据或信息后，我们对事件或假设的更新信念或估计。后验概率通常以$P(A|B)$表示，其中 A 是事件或假设，B 是新的数据或证据。贝叶斯定理告诉我们如何从先验概率和似然性（新数据在给定假设下的概率分布）计算后验概率，即在已知新信息的情况下，事件或假设的概率分布发生了什么变化。

马尔科夫链的定义、相关性质和公式

马尔科夫链（Markov Chain）是一个在时间序列中状态随机变化的数学模型，其中状态的转移仅依赖于前一个状态，而与过去的状态序列无关。它通常由以下几个要素定义：

1. **状态空间（State Space）：**马尔科夫链包含一组可能的状态，这些状态构成了状态空间，通常用符号 S 表示。
2. **转移概率（Transition Probability）：**对于任意两个状态 i 和 j，转移概率 P_{ij} 表示在当前状态为 i 的情况下，下一个状态为 j 的概率。这些概率构成了状态转移矩阵，通常用 P 表示，其中 P_{ij} 是第 i 行第 j 列的元素。
3. **初始分布（Initial Distribution）：**初始分布 π 表示系统在时间 $t = 0$ 时各个状态的概率分布。

相关性质：

1. **马尔科夫性质（Markov Property）：**马尔科夫链的核心特征是马尔科夫性质，即状态的转移只与当前状态相关，与过去状态无关。
2. **有限状态或无限状态：**马尔科夫链可以是有限状态的，即状态空间有限，也可以是无限状态的。
3. **时间齐次性（Time-Homogeneous）：**如果马尔科夫链的转移概率在时间上保持不变，即 P_{ij} 在不同时间步骤 t 之间保持不变，那么称它是时间齐次的。

马尔科夫链的状态转移概率公式：

马尔科夫链的状态转移概率公式描述了在给定当前状态的情况下，下一个状态的概率。对于有限状态空间的时间齐次马尔科夫链，状态转移概率可以表示为：$P(X_{t+1} = j|X_t = i) = P_{ij}$

这里，X_t 表示在时间 t 时的状态，P_{ij} 表示在状态 i 到状态 j 的转移概率。

马尔科夫链的状态序列：

马尔科夫链的状态序列是根据状态转移概率生成的随机序列，通常用 $X_0, X_1, X_2, \ldots$ 表示，其中 X_t 表示在时间 t 时的状态。这个序列可以用马尔科夫链的初始分布和状态转移概率来模拟或推断。

方差、标准差、协方差、协方差矩阵的定义、协方差与方差的关系

1. **方差（Variance）**：
   方差是一组数据的分散程度的度量，它衡量了数据集合中各个数据点与数据集合均值之间的差异程度。方差越大，数据点越分散。方差的数学定义如下：

   对于一组包含 n 个数据点的数据集 $x_1, x_2, \ldots, x_n$，其方差表示为 $Var(X)$，计算公式如下：

   $Var(X) = \Sigma[(x_i - \mu)^2]/n$

   其中，x_i 是数据点，μ 是数据集合的均值，n 是数据点的数量。
2. **标准差（Standard Deviation）**：
   标准差是方差的平方根，它用于测量数据集合的分散程度，以与原始数据的单位保持一致。标准差的计算公式如下：

   $StdDev(X) = \sqrt{Var(X)}$
3. **协方差（Covariance）**：
   协方差用于描述两个随机变量之间的线性关系。如果协方差为正，表示两个变量倾向于一起增加或减少；如果协方差为负，表示一个变量增加时，另一个变量倾向于减少。协方差的数学定义如下：

   对于两个随机变量 X 和 Y，其协方差表示为 $Cov(X, Y)$，计算公式如下：

   $Cov(X, Y) = \Sigma[(x_i - \mu_x) * (y_i - \mu_y)]/n$

   其中，x_i 和 y_i 分别是 X 和 Y 的观测值，μ_x 和 μ_y 分别是 X 和 Y 的均值，n 是观测值的数量。
4. **协方差矩阵（Covariance Matrix）**：
   协方差矩阵是一个方阵，用于描述多维数据中不同变量之间的协方差关系。对于包含多个随机变量的数据集，协方差矩阵将显示它们之间的协方差。协方差矩阵的对角线元素是各个变量的方差，非对角线元素是各个变量之间的协方差。

协方差与方差的关系：
协方差是方差的一种特殊情况。当计算协方差时，如果两个变量是同一个变量，即 X 和 Y 相等，那么协方差就等于该变量的方差。因此，方差是协方差的一种情况，即两个变量完全相同时的协方差。

协方差的计算公式中包括了两个变量之间的差值和均值，而方差的计算公式只包括了一个变量的差值和均值。因此，方差是协方差的一种特殊情况，用于描述单个变量的分散程度，而协方差用于描述两个变量之间的关系。

连续性随机变量的分布函数是不是处处可导

不一定。连续性随机变量的分布函数（也称为累积分布函数，Cumulative Distribution Function，CDF）不一定处处可导，其可导性取决于随机变量的概率密度函数（Probability Density Function，PDF）的性质。

具体来说，一个连续性随机变量的分布函数在某一点处可导，当且仅当该点对应的概率密度函数是连续可导的。这意味着如果概率密度函数在某点处有不连续点（例如，跳跃或尖峰），那么分布函数在该点处将不可导。

一些连续性随机变量的分布函数是处处可导的，例如正态分布（高斯分布）。对于这些分布，其概率密度函数在整个定义域上都是连续可导的，因此分布函数也是处处可导的。

然而，还有许多其他分布，如某些混合分布或分段函数，它们的概率密度函数可能在某些点上是不连续的，导致分布函数在这些点上不可导。这种情况下，分布函数可能会具有跳跃或垂直的部分，对应于概率密度函数的不连续点。

因此，连续性随机变量的分布函数的可导性取决于随机变量的概率密度函数的性质。通常，在概率论和统计学中，我们关心的是分布函数和概率密度函数的性质，以便更好地理解和分析随机变量的行为。

条件概率、联合概率、边缘概率的定义和区别

1. **条件概率（Conditional Probability）**：
   条件概率是指在已知某一事件发生的条件下，另一事件发生的概率。它表示为 $P(A|B)$，读作 "在事件 B 条件下事件 A 发生的概率"。条件概率的计算公式如下：$P(A|B) = P(A \cap B)/P(B)$

   其中，$P(A \cap B)$ 表示事件 A 和事件 B 同时发生的概率，$P(B)$ 是事件 B 发生的概率。条件概率允许我们在已知一些信息的情况下，重新评估其他事件的概率。

2. **联合概率（Joint Probability）**：
   联合概率是指多个事件同时发生的概率。对于两个事件 A 和 B，联合概率表示为 $P(A \cap B)$，它表示事件 A 和事件 B 同时发生的概率。对于多个事件，联合概率可以类似地扩展。联合概率可以用来描述多个事件之间的关系。

3. **边缘概率（Marginal Probability）**：
   边缘概率是指在联合概率的基础上计算单个事件的概率，而不考虑其他事件的影响。对于事件 A，它的边缘概率表示为 $P(A)$，通常是通过对联合概率 $P(A \cap B)$ 或 $P(A \cap B \cap C)$ 等等，对其他事件的所有可能情况进行求和或积分来计算的。

如果在二维随机变量(X,Y)的联合分布函数$F(x,y)$中令$y\to\infty$，由于$\{Y<\infty\}$为必然事件，故可得

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

这是由(X,Y)的联合分布函数$F(x,y)$求得的X的分布函数，被称为X的**边际分布**，记为

$$F_X(x)=F(x,\infty). \qquad (3.2.1)$$

类似地，在$F(x,y)$中令$x\to\infty$，可得Y的**边际分布**

$$F_Y(y)=F(\infty,y). \qquad (3.2.2)$$

区别：

- 条件概率考虑了一个事件在另一个事件已经发生的情况下的概率，它是在已知条件下的概率。
- 联合概率是多个事件同时发生的概率，它描述了事件之间的联合关系。
- 边缘概率是联合概率的一部分，它表示单个事件的概率，不考虑其他事件的影响。在计算边缘概率时，我们通常对联合概率进行求和或积分，以获得单个事件的概率。

这些概率之间的关系可以通过条件概率和边缘概率的关系来表示：

$P(A \cap B) = P(A|B) * P(B)$

$P(A) = \Sigma[P(A \cap B_i)]$，其中 B_i 表示多个事件 B 的可能取值。

概率分布函数和概率密度函数的定义和区别

1. **概率分布函数（分布函数）**：
   - **定义**：概率分布函数是一个函数，用于描述一个随机变量 X 取某个值以下的概率。它通常表示为 $F(x)$ 或 $P(X \leq x)$。
   - **性质**：概率分布函数的值在 $[0, 1]$ 之间，且满足以下性质：
     - $F(x)$ 是非递减的，即对于任意的 $a < b$，有 $F(a) \leq F(b)$。
     - $F(x)$ 在负无穷大处趋近于 0，且在正无穷大处趋近于 1。
   - **用途**：概率分布函数用于计算随机变量 X 小于或等于某个特定值的概率，即 $P(X \leq x)$。

2. **概率密度函数（密度函数）**：
   - **定义**：概率密度函数是一个函数，用于描述连续型随机变量的概率分布。它通常表示为 $f(x)$，并且满足以下性质：
     - $f(x) \geq 0$，即密度函数的值非负。

        - 在整个定义域内的积分等于 1，即 $\int[f(x)]dx = 1$。
    - **性质**：密度函数表示随机变量在某一区间内的概率密度，而不是概率本身。具体概率由对密度函数的积分来计算。
    - **用途**：概率密度函数用于描述连续型随机变量的概率分布，通过对密度函数进行积分可以计算随机变量落在某个区间内的概率。

区别：

- **类型**：概率分布函数通常用于描述离散型和连续型随机变量的概率分布，而概率密度函数主要用于描述连续型随机变量的概率分布。
- **性质**：概率分布函数是一个非递减函数，其值介于 $[0, 1]$ 之间，用于计算随机变量小于或等于某个值的概率。概率密度函数是非负的函数，其积分等于 1，用于描述随机变量在某一点附近的概率密度。
- **计算概率**：对于概率分布函数，概率可以直接从分布函数中读取。对于概率密度函数，概率需要通过对密度函数的积分来计算。

二项分布、正态分布、指数分布、泊松分布的定义、实际应用与联系

1. **二项分布（Binomial Distribution）**：
    - **定义**：二项分布描述了一系列独立的伯努利试验，每次试验都有两种可能的结果，成功和失败。二项分布表示在 n 次试验中成功发生的次数，其中每次试验成功的概率为 p。其概率质量函数（PMF）为二项式系数的形式。
    - **实际应用**：常用于模拟二元事件的多次重复，如硬币投掷、产品质量检验、投票结果分析等。
2. **正态分布（Normal Distribution）**：
    - **定义**：正态分布是连续型概率分布，具有钟形曲线形状。其均值和标准差决定了分布的中心位置和分散程度。正态分布是许多自然现象和随机过程的模型，具有很多性质，如68-95-99.7法则。
    - **实际应用**：广泛用于描述身高、体重、考试成绩、温度、金融市场波动等连续型随机变量的分布。
3. **指数分布（Exponential Distribution）**：
    - **定义**：指数分布是连续型概率分布，通常用于描述随机事件的等待时间或间隔时间。其参数 λ 表示单位时间内事件发生的平均率。指数分布的概率密度函数呈指数衰减形式。
    - **实际应用**：常用于建模随机事件的等待时间，如设备的故障间隔时间、电话呼叫之间的时间、到达顾客的时间间隔等。
4. **泊松分布（Poisson Distribution）**：
    - **定义**：泊松分布用于描述在固定时间或空间区间内某事件的发生次数，其参数 λ 表示单位时间或单位空间区间内事件的平均发生率。泊松分布是二项分布在试验次数很大、成功概率很小的极限情况。
    - **实际应用**：常用于建模稀有事件的发生，如电话呼叫中心的呼叫次数、车辆交通事故的数量、错误检测等。

联系：

- 正态分布可以用中心极限定理来近似描述一组独立随机变量的和或平均值的分布，因此在实际应用中与其他分布（如二项分布）有关。
- 当二项分布中的试验次数 n 很大且成功概率 p 很小时，可以用泊松分布来近似描述二项分布。
- 指数分布可以看作泊松分布的连续型版本，因为它们都描述了事件发生的等待时间。当事件发生的次数很大时，泊松分布的等待时间可以近似为指数分布。

有哪些常用的概率分布

表 2.5.1　常用概率分布及其数学期望和方差

分　布	分布列 p_k 或分布密度 $p(x)$	期　望	方　差
0-1 分布	$p_k=p^k(1-p)^{1-k}$，　$k=0,1$	p	$p(1-p)$
二项分布 $b(n,p)$	$p_k=\binom{n}{k}p^k(1-p)^{n-k}$，　$k=0,1,\cdots,n$	np	$np(1-p)$
泊松分布 $P(\lambda)$	$p_k=\dfrac{\lambda^k}{k!}e^{-\lambda}$，　$k=0,1,\cdots$	λ	λ
超几何分布 $h(n,N,M)$	$p_k=\dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$，　$k=0,1,\cdots,r$，$r=\min\{M,n\}$	$n\dfrac{M}{N}$	$\dfrac{nM(N-M)(N-n)}{N^2(N-1)}$
几何分布 $Ge(p)$	$p_k=(1-p)^{k-1}p$，　$k=1,2,\cdots$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$
负二项分布 $Nb(r,p)$	$p_k=\binom{k-1}{r-1}(1-p)^{k-r}p^r$，　$k=r,r+1,\cdots$	$\dfrac{r}{p}$	$\dfrac{r(1-p)}{p^2}$
正态分布 $N(\mu,\sigma^2)$	$p(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$，　$-\infty<x<\infty$	μ	σ^2
均匀分布 $U(a,b)$	$p(x)=\dfrac{1}{b-a}$，　$a<x<b$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$

续表

分　布	分布列 p_k 或分布密度 $p(x)$	期　望	方　差
指数分布 $Exp(\lambda)$	$p(x)=\lambda e^{-\lambda x}$，　$x\geqslant 0$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
伽马分布 $Ga(\alpha,\lambda)$	$p(x)=\dfrac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$，　$x\geqslant 0$	$\dfrac{\alpha}{\lambda}$	$\dfrac{\alpha}{\lambda^2}$
$\chi^2(n)$ 分布	$p(x)=\dfrac{x^{n/2-1}e^{-x/2}}{\Gamma(n/2)2^{n/2}}$，　$x\geqslant 0$	n	$2n$
贝塔分布 $Be(a,b)$	$p(x)=\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$，　$0<x<1$	$\dfrac{a}{a+b}$	$\dfrac{ab}{(a+b)^2(a+b+1)}$
对数正态分布 $LN(\mu,\sigma^2)$	$p(x)=\dfrac{1}{\sqrt{2\pi}\sigma x}\exp\left\{-\dfrac{(\ln x-\mu)^2}{2\sigma^2}\right\}$，$x>0$	$e^{\mu+\sigma^2/2}$	$e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$
柯西分布 $\mathrm{Cau}(\mu,\lambda)$	$p(x)=\dfrac{1}{\pi}\dfrac{\lambda}{\lambda^2+(x-\mu)^2}$，$-\infty<x<\infty$	不存在	不存在
韦布尔分布	$p(x)=F'(x)$，$F(x)=1-\exp\left\{-\left(\dfrac{x}{\eta}\right)^m\right\}$，$x>0$	$\eta\Gamma\left(1+\dfrac{1}{m}\right)$	$\eta^2\left[\Gamma\left(1+\dfrac{2}{m}\right)-\Gamma^2\left(1+\dfrac{1}{m}\right)\right]$

注：表中仅列出各分布密度函数的非零区域.

无偏估计和有偏估计的定义

1. **无偏估计（Unbiased Estimation）**：
   无偏估计是一种估计方法，其估计值的期望值等于被估计参数的真实值。换句话说，如果使用无偏估计方法多次进行估计，估计值的平均值将等于真实值。数学上，对于参数 θ 和无偏估计器 T(X)，满足以下条件：$E[T(X)] = \theta$

   这表示无偏估计器的期望值为真实参数值。无偏估计在统计学中很有价值，因为它们不会引入估计偏差，即估计值不会有系统性的高估或低估。

2. **有偏估计（Biased Estimation）**：
   有偏估计是一种估计方法，其估计值的期望值与被估计参数的真实值不相等。换句话说，有偏估计引入了估计偏差，可能高估或低估真实参数值。数学上，对于参数 θ 和有偏估计器 T(X)，满足以下条件：$E[T(X)] \neq \theta$

   这表示有偏估计器的期望值不等于真实参数值。尽管有偏估计可能不会准确估计参数的真实值，但它们在某些情况下仍然有用，因为它们可能具有其他有用的性质，如较低的方差或更简单的计算形式。有时，有偏估计也可以通过一些修正方法来转化为无偏估计。

定义 6.1.1 设 $x_1, x_2, \cdots, x_n$ 是来自总体的一个样本，用于估计未知参数 θ 的统计量 $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$ 称为 θ 的估计量，或称为 θ 的**点估计**，简称**估计**.

在这里如何构造统计量 $\hat{\theta}$ 并没有明确的规定，只要它满足一定的合理性即可.最常见的合理性要求是所谓的无偏性.

定义 6.1.2 设 $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$ 是 θ 的一个估计，θ 的参数空间为 Θ，若对任意的 $\theta \in \Theta$，有

$$E_\theta(\hat{\theta}) = \theta, \tag{6.1.1}$$

则称 $\hat{\theta}$ 是 θ 的**无偏估计**，否则称为**有偏估计**.

最小二乘法的过程

1. **定义模型**：
   首先，你需要定义一个线性模型，通常表示为：

   $Y = \beta_0 + \beta_1 X + \varepsilon$

   - Y 是因变量（要预测的变量）。
   - X 是自变量（已知的输入或特征）。
   - β_0 和 β_1 是要估计的参数，分别表示截距和斜率。
   - ε 是误差项，表示模型无法完全解释的随机噪声。
2. **收集数据**：
   收集包含因变量 Y 和自变量 X 的数据样本。通常，你会有多个数据点，以便进行估计。
3. **拟合模型**：
   使用最小二乘法，你的目标是找到最佳的参数估计值 β_0 和 β_1，以使模型的预测值与实际观测值之间的误差平方和最小化。

   具体来说，最小二乘法的目标是最小化残差平方和（Sum of Squared Residuals，SSR），也就是误差的平方和：$SSR = \Sigma(y_i - \hat{y}_i)^2$

   其中，y_i 是观测到的因变量值，$\hat{y}_i$ 是模型给出的相应预测值，Σ 表示对所有数据点求和。
4. **计算参数估计**：
   通过对 SSR 对参数 β_0 和 β_1 求偏导数，并将其置为零，可以得到参数估计的闭合解。这些估计值是最小二乘法的估计结果。
   - 估计 β_0：通过对 SSR 对 β_0 求偏导数并令其等于零，解出 β_0 的估计值。

- 估计 β_1：通过对 SSR 对 β_1 求偏导数并令其等于零，解出 β_1 的估计值。

这些估计值将使 SSR 最小化。

5. **模型评估**：
   一旦获得参数估计，你可以使用这些估计值来构建最佳拟合线性模型。然后，你可以评估模型的质量，例如通过检查残差、计算拟合的 R^2 值等。

一般采用最小二乘方法估计模型(8.4.5)中的 β_0,β_1.令

$$Q(\beta_0,\beta_1)=\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2,$$

$\hat{\beta}_0,\hat{\beta}_1$ 应该满足

$$Q(\hat{\beta}_0,\hat{\beta}_1)=\min_{\beta_0,\beta_1}Q(\beta_0,\beta_1),$$

这样得到的$\hat{\beta}_0,\hat{\beta}_1$ 称为 β_0,β_1 的**最小二乘估计**,记为 LSE.

由于 $Q\geqslant 0$,且对 β_0,β_1 的导数存在,因此最小二乘估计可以通过求偏导数并命其为 0 而得到

$$\begin{cases}\dfrac{\partial Q}{\partial\beta_0}=-2\sum\limits_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)=0,\\ \dfrac{\partial Q}{\partial\beta_1}=-2\sum\limits_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)x_i=0.\end{cases}\tag{8.4.7}$$

这组方程称为**正规方程组**,经过整理,可得

$$\begin{cases}n\beta_0+n\bar{x}\beta_1=n\bar{y},\\ n\bar{x}\beta_0+\sum x_i^2\beta_1=\sum x_iy_i,\end{cases}\tag{8.4.8}$$

(今后凡是不作说明“$\sum$”都表示“$\sum\limits_{i=1}^{n}$”.)记

$$\bar{x}=\frac{1}{n}\sum x_i,\quad \bar{y}=\frac{1}{n}\sum y_i,$$

$$l_{xy}=\sum(x_i-\bar{x})(y_i-\bar{y})=\sum x_iy_i-n\bar{x}\cdot\bar{y}=\sum x_iy_i-\frac{1}{n}\sum x_i\sum y_i,$$

$$l_{xx}=\sum(x_i-\bar{x})^2=\sum x_i^2-n\bar{x}^2=\sum x_i^2-\frac{1}{n}(\sum x_i)^2,$$

$$l_{yy}=\sum(y_i-\bar{y})^2=\sum y_i^2-n\bar{y}^2=\sum y_i^2-\frac{1}{n}(\sum y_i)^2.$$

解(8.4.8)可得

$$\begin{cases}\hat{\beta}_1=\dfrac{l_{xy}}{l_{xx}},\\ \hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}.\end{cases}\tag{8.4.9}$$

白噪声是什么，有什么特点

白噪声（White Noise）是一种随机信号或随机过程，其主要特点是具有以下性质：

1. **随机性**：白噪声是完全随机的信号，没有明显的可预测性。每个时刻的值是独立、不相关的，并且没有任何可辨识的模式。
2. **恒定功率**：在频谱上，白噪声在所有频率上具有近似相等的功率。这意味着它的能量在不同频率上均匀分布，没有频率偏向。
3. **均值为零**：白噪声的均值等于零，即其期望值为零。这表示白噪声在长期内没有明显的趋势或漂移。
4. **有限带宽**：在实际应用中，白噪声通常有一个有限的频率范围，因为无限频带的白噪声在现实中是不可实现的。这种有限带宽的白噪声称为"有限带宽白噪声"。