

## 第 1 章 随机事件及其概率

<p>(1) 排列</p> <p>组合公式</p>	$P_m^n = \frac{m!}{(m-n)!}$ <p>从 <math>m</math> 个人中挑出 <math>n</math> 个人进行排列的可能数。</p> $C_m^n = \frac{m!}{n!(m-n)!}$ <p>从 <math>m</math> 个人中挑出 <math>n</math> 个人进行组合的可能数。</p>
<p>(2) 加法和乘法原理</p>	<p><b>加法原理</b> (两种方法均能完成此事): <math>m+n</math></p> <p>某件事由两种方法来完成, 第一种方法可由 <math>m</math> 种方法完成, 第二种方法可由 <math>n</math> 种方法来完成, 则这件事可由 <math>m+n</math> 种方法来完成。</p> <p><b>乘法原理</b> (两个步骤分别不能完成这件事): <math>m \times n</math></p> <p>某件事由两个步骤来完成, 第一个步骤可由 <math>m</math> 种方法完成, 第二个步骤可由 <math>n</math> 种方法来完成, 则这件事可由 <math>m \times n</math> 种方法来完成。</p>
<p>(3) 一些常见排列</p>	<p>重复排列和非重复排列 (有序)</p> <p>对立事件 (至少有一个)</p> <p>顺序问题</p>
<p>(4) 随机试验和随机事件</p>	<p>如果一个试验在相同条件下可以重复进行, 而每次试验的可能结果不止一个, 但在进行一次试验之前却不能断言它出现哪个结果, 则称这种试验为<b>随机试验</b>。</p> <p>试验的可能结果称为<b>随机事件</b>。</p>
<p>(5) 基本事件、样本空间和事件</p>	<p>在一个试验下, 不管事件有多少个, 总可以从其中找出这样一组事件, 它具有如下性质:</p> <ul style="list-style-type: none"> <li>✓ 每进行一次试验, 必须发生且只能发生这一组中的一个事件;</li> <li>✓ 任何事件, 都是由这一组中的部分事件组成的。</li> </ul> <p>这样一组事件中的每一个事件称为<b>基本事件</b>, 用 <math>\omega</math> 来表示。</p> <p>基本事件的全体, 称为试验的<b>样本空间</b>, 用 <math>\Omega</math> 表示。</p> <p>一个<b>事件</b>就是由 <math>\Omega</math> 中的部分点 (基本事件 <math>\omega</math>) 组成的集合。通常用大写字母 <math>A, B, C, \dots</math> 表示事件, 它们是 <math>\Omega</math> 的子集。</p>

	<p><math>\Omega</math> 为必然事件, <math>\emptyset</math> 为不可能事件。</p> <p>不可能事件 (<math>\emptyset</math>) 的概率为零, 而概率为零的事件不一定是不可能事件; 同理, 必然事件 (<math>\Omega</math>) 的概率为 1, 而概率为 1 的事件也不一定是必然事件。</p>
(6) 事件的关系与运算	<p>①关系:</p> <ul style="list-style-type: none"> <li>✓ 如果事件 <math>A</math> 的组成部分也是事件 <math>B</math> 的组成部分, (<math>A</math> 发生必有事件 <math>B</math> 发生):  <math>A \subset B</math></li> <li>✓ 如果同时有 <math>A \subset B, B \supset A</math>, 则称事件 <math>A</math> 与事件 <math>B</math> 等价, 或称 <math>A</math> 等于 <math>B</math>: <math>A=B</math>.</li> <li>✓ <math>A, B</math> 中至少有一个发生的事件: <math>A \cup B</math>, 或者 <math>A+B</math>.</li> <li>✓ 属于 <math>A</math> 而不属于 <math>B</math> 的部分所构成的事件, 称为 <math>A</math> 与 <math>B</math> 的差, 记为 <math>A-B</math>, 也可表示为 <math>A-AB</math> 或者 <math>A\bar{B}</math>, 它表示 <math>A</math> 发生而 <math>B</math> 不发生的事件。</li> <li>✓ <math>A, B</math> 同时发生: <math>A \cap B</math>, 或者 <math>AB</math>。 <math>A \cap B = \emptyset</math>, 则表示 <math>A</math> 与 <math>B</math> 不可能同时发生, 称事件 <math>A</math> 与事件 <math>B</math> 互不相容或者互斥。基本事件是互不相容的。</li> <li>✓ <math>\Omega-A</math> 称为事件 <math>A</math> 的逆事件, 或称 <math>A</math> 的对立事件, 记为 <math>\bar{A}</math>。它表示 <math>A</math> 不发生的事件。互斥未必对立。</li> </ul> <p>②运算:</p> <p>交换律: <math>A \cup B = B \cup A \quad A \cap B = B \cap A</math></p> <p>结合率: <math>A(BC) = (AB)C \quad A \cup (B \cap C) = (A \cup B) \cap C</math></p> <p>分配率: <math>(AB) \cup C = (A \cup C) \cap (B \cup C) \quad (A \cup B) \cap C = (AC) \cup (BC)</math></p> <p>德摩根率: <math>\overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \bar{A}_i \quad \overline{A \cup B} = \bar{A} \cap \bar{B}, \quad \overline{A \cap B} = \bar{A} \cup \bar{B}</math></p>
(7) 概率的公理化定义	<p>设 <math>\Omega</math> 为样本空间, <math>A</math> 为事件, 对每一个事件 <math>A</math> 都有一个实数 <math>P(A)</math>, 若满足下列三个条件:</p> <p>1° <math>0 \leq P(A) \leq 1</math>, 非负性</p> <p>2° <math>P(\Omega) = 1</math>, 规范性</p>

	<p>3° 对于两两互不相容的事件 <math>A_1, A_2, \dots</math> 有</p> $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ <p>常称为<b>可列（完全）可加性</b>。</p> <p>则称 <math>P(A)</math> 为事件 <math>A</math> 的<b>概率</b>。</p>
(8) 古典 概型	<p>1° <math>\Omega = \{\omega_1, \omega_2 \cdots \omega_n\}</math>,</p> <p>2° <math>P(\omega_1) = P(\omega_2) = \cdots P(\omega_n) = \frac{1}{n}</math>。</p> <p>设任一事件 <math>A</math>，它是由 <math>\omega_1, \omega_2 \cdots \omega_m</math> 组成的，则有</p> $P(A) = \{(\omega_1) \cup (\omega_2) \cup \cdots \cup (\omega_m)\} = P(\omega_1) + P(\omega_2) + \cdots + P(\omega_m)$ $= \frac{m}{n} = \frac{A \text{ 所包含的基本事件数}}{\text{基本事件总数}}$
(9) 几何 概型	<p>若随机试验的结果为无限不可数并且每个结果出现的可能性均匀，同时样本空间中的每一个基本事件可以使用一个有界区域来描述，则称此随机试验为<b>几何概型</b>。</p> <p>对任一事件 <math>A</math>，<math>P(A) = \frac{L(A)}{L(\Omega)}</math>。其中 <math>L</math> 为几何度量（长度、面积、体积）。</p>
(10) 加 法公式	$P(A+B) = P(A) + P(B) - P(AB)$ <p>当 <math>P(AB) = 0</math> 时，<math>P(A+B) = P(A) + P(B)</math></p>
(11) 减 法公式	$P(A-B) = P(A) - P(AB)$ <p>当 <math>B \subset A</math> 时，<math>P(A-B) = P(A) - P(B)</math></p> <p>当 <math>A = \Omega</math> 时，<math>P(\bar{B}) = 1 - P(B)</math></p>
(12) 条 件概率	<p>定义 设 <math>A, B</math> 是两个事件，且 <math>P(A) &gt; 0</math>，则称 <math>\frac{P(AB)}{P(A)}</math> 为事件 <math>A</math> 发生条件下，事件 <math>B</math> 发生的<b>条件概率</b>，记为 <math>P(B/A) = \frac{P(AB)}{P(A)}</math>。</p> <p>条件概率是概率的一种，所有概率的性质都适合于条件概率。</p>

	例如 $P(\Omega/B)=1 \Rightarrow P(\bar{B}/A)=1-P(B/A)$
(13) 乘法公式	<p>乘法公式: <math>P(AB) = P(A)P(B/A)</math></p> <p>更一般地, 对事件 <math>A_1, A_2, \dots, A_n</math>, 若 <math>P(A_1A_2\dots A_{n-1}) &gt; 0</math>, 则有</p> $P(A_1A_2\dots A_n) = P(A_1)P(A_2 A_1)P(A_3 A_1A_2)\dots P(A_n A_1A_2\dots A_{n-1})。$
(14) 独立性	<p>①两个事件的独立性</p> <ul style="list-style-type: none"> <li>✓ 设事件 <math>A、B</math> 满足 <math>P(AB) = P(A)P(B)</math>, 则称事件 <math>A、B</math> 是相互独立的。</li> <li>✓ 若事件 <math>A、B</math> 相互独立, 且 <math>P(A) &gt; 0</math>, 则有</li> </ul> $P(B A) = \frac{P(AB)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B)$ <ul style="list-style-type: none"> <li>✓ 若事件 <math>A、B</math> 相互独立, 则可得到 <math>\bar{A}</math> 与 <math>B</math>、<math>A</math> 与 <math>\bar{B}</math>、<math>\bar{A}</math> 与 <math>\bar{B}</math> 也都相互独立。</li> <li>✓ 必然事件 <math>\Omega</math> 和不可能事件 <math>\emptyset</math> 与任何事件都相互独立。</li> <li>✓ <math>\emptyset</math> 与任何事件都互斥。</li> </ul> <p>②多个事件的独立性</p> <p>设 <math>A, B, C</math> 是三个事件, 如果满足两两独立的条件,</p> $P(AB)=P(A)P(B); P(BC)=P(B)P(C); P(CA)=P(C)P(A)$ <p>并且同时满足 <math>P(ABC)=P(A)P(B)P(C)</math></p> <p>那么 <math>A、B、C</math> <b>相互独立</b>。</p> <p>对于 <math>n</math> 个事件类似。</p>
(15) 全概率公式	<p>设事件 <math>B_1, B_2, \dots, B_n</math> 满足</p> <ol style="list-style-type: none"> <li><math>B_1, B_2, \dots, B_n</math> 两两互不相容, <math>P(B_i) &gt; 0 (i = 1, 2, \dots, n)</math>,</li> <li><math>A \subset \bigcup_{i=1}^n B_i</math>,</li> </ol> <p>则有</p> $P(A) = P(B_1)P(A B_1) + P(B_2)P(A B_2) + \dots + P(B_n)P(A B_n)。$

<p>(16) 贝叶斯公式</p>	<p>设事件 <math>B_1, B_2, \dots, B_n</math> 及 <math>A</math> 满足</p> <p>1° <math>B_1, B_2, \dots, B_n</math> 两两互不相容, <math>P(B_i) &gt; 0, i = 1, 2, \dots, n,</math></p> <p>2° <math>A \subset \bigcup_{i=1}^n B_i, P(A) &gt; 0,</math></p> <p>则</p> $P(B_i / A) = \frac{P(B_i)P(A / B_i)}{\sum_{j=1}^n P(B_j)P(A / B_j)}, i=1, 2, \dots, n.$ <p>此公式即为<b>贝叶斯公式</b>。</p> <p><math>P(B_i), (i=1, 2, \dots, n),</math> 通常叫<b>先验概率</b>。 <math>P(B_i / A), (i=1, 2, \dots, n),</math> 通常称为<b>后验概率</b>。贝叶斯公式反映了“因果”的概率规律, 并作出了“由果溯因”的推断。</p>
<p>(17) 伯努利概型</p>	<p>我们作了 <math>n</math> 次试验, 且满足</p> <ul style="list-style-type: none"> <li>✓ 每次试验只有两种可能结果, <math>A</math> 发生或 <math>A</math> 不发生;</li> <li>✓ <math>n</math> 次试验是重复进行的, 即 <math>A</math> 发生的概率每次均一样;</li> <li>✓ 每次试验是独立的, 即每次试验 <math>A</math> 发生与否与其他次试验 <math>A</math> 发生与否是互不影响的。</li> </ul> <p>这种试验称为<b>伯努利概型</b>, 或称为 <math>n</math> 重伯努利试验。</p> <p>用 <math>p</math> 表示每次试验 <math>A</math> 发生的概率, 则 <math>\bar{A}</math> 发生的概率为 <math>1 - p = q</math>, 用 <math>P_n(k)</math> 表示 <math>n</math> 重伯努利试验中 <math>A</math> 出现 <math>k(0 \leq k \leq n)</math> 次的概率,</p> $P_n(k) = C_n^k p^k q^{n-k}, k = 0, 1, 2, \dots, n.$

## 第二章 随机变量及其分布

<p>(1) 离散型随机变量的分布律</p>	<p>设离散型随机变量 <math>X</math> 的可能取值为 <math>X_k(k=1,2,\dots)</math>且取各个值的概率，即事件<math>(X=X_k)</math>的概率为 <math>P(X=x_k)=p_k, \quad k=1,2,\dots,</math></p> <p>则称上式为<b>离散型随机变量 <math>X</math> 的概率分布或分布律</b>。有时也用分布列的形式给出：</p> $\frac{X}{P(X=x_k)} \mid \frac{x_1, x_2, \dots, x_k, \dots}{p_1, p_2, \dots, p_k, \dots}.$ <p>显然分布律应满足下列条件：</p> <p>(1) <math>p_k \geq 0, \quad k=1,2,\dots, \quad (2) \sum_{k=1}^{\infty} p_k = 1</math>。</p>
<p>(2) 连续型随机变量的分布密度</p>	<p>设 <math>F(x)</math> 是随机变量 <math>X</math> 的分布函数，若存在非负函数 <math>f(x)</math>，对任意实数 <math>x</math>，有</p> $F(x) = \int_{-\infty}^x f(x)dx,$ <p>则称 <math>X</math> 为<b>连续型随机变量</b>。 <math>f(x)</math> 称为 <math>X</math> 的概率密度函数或密度函数，简称<b>概率密度</b>。</p> <p>密度函数具有下面 4 个性质：</p> <p>1° <math>f(x) \geq 0</math>;</p> <p>2° <math>\int_{-\infty}^{+\infty} f(x)dx = 1</math>;</p> <p>3° 对于任意实数 <math>x_1, x_2(x_1 \leq x_2)</math>, <math>P\{x_1 &lt; X \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx</math>;</p> <p>4° 若 <math>f(x)</math> 在点 <math>x</math> 处连续，则有 <math>F'(x) = f(x)</math>。</p>
<p>(3) 离散与连续型随机变量的关系</p>	<p><math>P(X=x) \approx P(x &lt; X \leq x+dx) \approx f(x)dx</math></p> <p>积分元 <math>f(x)dx</math> 在连续型随机变量理论中所起的作用与 <math>P(X=x_k)=p_k</math> 在离散型随机变量理论中所起的作用相类似。</p>

<p>(4) 分布函数</p>	<p>设 <math>X</math> 为随机变量, <math>x</math> 是任意实数, 则函数</p> $F(x) = P(X \leq x)$ <p>称为<b>随机变量 <math>X</math> 的分布函数</b>, 本质上是一个累积函数。</p> <p><math>P(a &lt; X \leq b) = F(b) - F(a)</math> 可以得到 <math>X</math> 落入区间 <math>(a, b]</math> 的概率。分布函数 <math>F(x)</math> 表示随机变量落入区间 <math>(-\infty, x]</math> 内的概率。</p> <p>分布函数具有如下<b>性质</b>:</p> <p>1° <math>0 \leq F(x) \leq 1, \quad -\infty &lt; x &lt; +\infty;</math></p> <p>2° <math>F(x)</math> 是单调不减的函数, 即 <math>x_1 &lt; x_2</math> 时, 有 <math>F(x_1) \leq F(x_2);</math></p> <p>3° <math>F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0, \quad F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1;</math></p> <p>4° <math>F(x+0) = F(x),</math> 即 <math>F(x)</math> 是右连续的;</p> <p>5° <math>P(X = x) = F(x) - F(x-0)。</math></p> <p>对于离散型随机变量, <math>F(x) = \sum_{x_k \leq x} p_k;</math></p> <p>对于连续型随机变量, <math>F(x) = \int_{-\infty}^x f(x)dx。</math></p>	
<p>(5) 八大分布</p>	<p>0-1 分布</p>	<p><math>P(X=1)=p, P(X=0)=q</math></p> <hr/> <p>二项分布</p> <p>在 <math>n</math> 重贝努里试验中, 设事件 <math>A</math> 发生的概率为 <math>p</math>。事件 <math>A</math> 发生的次数是随机变量, 设为 <math>X</math>, 则 <math>X</math> 可能取值为 <math>0, 1, 2, \dots, n。</math></p> <p><math>P(X = k) = P_n(k) = C_n^k p^k q^{n-k}, \quad \text{其中 } q = 1 - p, 0 &lt; p &lt; 1, k = 0, 1, 2, \dots, n,</math></p> <p>则称随机变量 <math>X</math> 服从参数为 <math>n, p</math> 的<b>二项分布</b>。记为 <math>X \sim B(n, p)。</math></p> <p>当 <math>n = 1</math> 时, <math>P(X = k) = p^k q^{1-k}, \quad k = 0, 1,</math> 这就是 (0-1) 分布, 所以 (0-1) 分布是二项分布的特例。</p>

泊松分布	<p>设随机变量 <math>X</math> 的分布律为</p> $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda > 0, \quad k = 0, 1, 2, \dots,$ <p>则称随机变量 <math>X</math> 服从参数为 <math>\lambda</math> 的<b>泊松分布</b>, 记为 <math>X \sim \pi(\lambda)</math> 或者 <math>P(\lambda)</math>。</p> <p>泊松分布为二项分布的极限分布 (<math>np = \lambda, n \rightarrow \infty</math>)。</p>
超几何分布	$P(X = k) = \frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n}, \quad k = 0, 1, 2, \dots, l$ <p>随机变量 <math>X</math> 服从参数为 <math>n, N, M</math> 的<b>超几何分布</b>, 记为 <math>H(n, N, M)</math>。</p>
几何分布	$P(X = k) = q^{k-1} p, \quad k = 1, 2, 3, \dots,$ <p>其中 <math>p \geq 0, q = 1 - p</math>。</p> <p>随机变量 <math>X</math> 服从参数为 <math>p</math> 的<b>几何分布</b>, 记为 <math>G(p)</math>。</p>
均匀分布	<p>设随机变量 <math>X</math> 的值只落在 <math>[a, b]</math> 内, 其密度函数 <math>f(x)</math> 在 <math>[a, b]</math> 上为常数 <math>\frac{1}{b-a}</math>, 即</p> $f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{其他,} \end{cases}$ <p>则称随机变量 <math>X</math> 在 <math>[a, b]</math> 上服从<b>均匀分布</b>, 记为 <math>X \sim U(a, b)</math>。</p> <p>分布函数为</p> $F(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b. \end{cases}$ <p>当 <math>a \leq x_1 &lt; x_2 \leq b</math> 时, <math>X</math> 落在区间 <math>(x_1, x_2)</math> 内的概率为</p> $P(x_1 < X < x_2) = \frac{x_2 - x_1}{b - a}.$



	<p>指数分布</p> $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$ <p>其中 <math>\lambda &gt; 0</math> , 则称随机变量 X 服从参数为 <math>\lambda</math> 的<b>指数分布</b>。</p> <p>X 的分布函数为</p> $F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$ <p>记住积分公式:</p> $\int_0^{+\infty} x^n e^{-x} dx = n!$
--	--

	<p>正态分布</p>	<p>设随机变量 <math>X</math> 的密度函数为</p> $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty,$ <p>其中 <math>\mu</math>、<math>\sigma &gt; 0</math> 为常数, 则称随机变量 <math>X</math> 服从参数为 <math>\mu</math>、<math>\sigma</math> 的<b>正态分布或高斯 (Gauss) 分布</b>, 记为 <math>X \sim N(\mu, \sigma^2)</math>。</p> <p><math>f(x)</math> 具有如下性质:</p> <p>1° <math>f(x)</math> 的图形是关于 <math>x = \mu</math> 对称的;</p> <p>2° 当 <math>x = \mu</math> 时, <math>f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}</math> 为最大值;</p> <p>若 <math>X \sim N(\mu, \sigma^2)</math>, 则 <math>X</math> 的分布函数为</p> $F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt。$ <p>参数 <math>\mu = 0</math>、<math>\sigma = 1</math> 时的正态分布称为<b>标准正态分布</b>, 记为 <math>X \sim N(0,1)</math>, 其密度函数记为</p> $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < +\infty,$ <p>分布函数为</p> $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt。$ <p><math>\Phi(x)</math> 是不可求积函数, 其函数值, 已编制成表可供查用。</p> <p><math>\Phi(-x) = 1 - \Phi(x)</math> 且 <math>\Phi(0) = \frac{1}{2}</math>。</p> <p>如果 <math>X \sim N(\mu, \sigma^2)</math>, 则 <math>\frac{X - \mu}{\sigma} \sim N(0,1)</math>。</p> $P(x_1 < X \leq x_2) = \Phi\left(\frac{x_2 - \mu}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu}{\sigma}\right)。$
--	-------------	--

(6) 分位数	下分位表: $P(X \leq \mu_\alpha) = \alpha$ ; 上分位表: $P(X > \mu_\alpha) = \alpha$ 。	
(7) 函数分布	离散型	<p>已知 <math>X</math> 的分布列为</p> $\frac{X}{P(X = x_i)} \left  \begin{array}{c} x_1, x_2, \dots, x_n, \dots \\ p_1, p_2, \dots, p_n, \dots \end{array} \right. ,$ <p><math>Y = g(X)</math> 的分布列 (<math>y_i = g(x_i)</math> 互不相等) 如下:</p> $\frac{Y}{P(Y = y_i)} \left  \begin{array}{c} g(x_1), g(x_2), \dots, g(x_n), \dots \\ p_1, p_2, \dots, p_n, \dots \end{array} \right. ,$ <p>若有某些 <math>g(x_i)</math> 相等, 则应将对应的 <math>p_i</math> 相加作为 <math>g(x_i)</math> 的概率。</p>
	连续型	先利用 $X$ 的概率密度 $f_X(x)$ 写出 $Y$ 的分布函数 $F_Y(y) = P(g(X) \leq y)$ , 再利用变上下限积分的求导公式求出 $f_Y(y)$ 。

### 第三章 二维随机变量及其分布

(1) 联合分布	离散型	如果二维随机向量 $\xi$ ( $X$ , $Y$ ) 的所有可能取值为至多可列个有序对 $(x,y)$ , 则称 $\xi$ 为 <b>离散型随机量</b> 。																																				
		设 $\xi = (X, Y)$ 的所有可能取值为 $(x_i, y_j)(i, j = 1,2,\cdots)$ , 且事件 $\{\xi = (x_i, y_j)\}$ 的概率为 $p_{ij}$ ,称																																				
		$P\{(X,Y) = (x_i, y_j)\} = p_{ij}(i, j = 1,2,\cdots)$																																				
		为 $\xi = (X, Y)$ 的分布律或称为 $X$ 和 $Y$ 的 <b>联合分布律</b> 。联合分布有时也用下面的概率分布表来表示:																																				
		<table><tr><td><math>\begin{matrix} &amp; Y \\ X \backslash &amp; \end{matrix}</math></td><td><math>y_1</math></td><td><math>y_2</math></td><td><math>\cdots</math></td><td><math>y_j</math></td><td><math>\cdots</math></td></tr><tr><td><math>x_1</math></td><td><math>p_{11}</math></td><td><math>p_{12}</math></td><td><math>\cdots</math></td><td><math>p_{1j}</math></td><td><math>\cdots</math></td></tr><tr><td><math>x_2</math></td><td><math>p_{21}</math></td><td><math>p_{22}</math></td><td><math>\cdots</math></td><td><math>p_{2j}</math></td><td><math>\cdots</math></td></tr><tr><td><math>\vdots</math></td><td><math>\vdots</math></td><td><math>\vdots</math></td><td></td><td><math>\vdots</math></td><td><math>\vdots</math></td></tr><tr><td><math>x_i</math></td><td><math>p_{i1}</math></td><td></td><td><math>\cdots</math></td><td><math>p_{ij}</math></td><td><math>\cdots</math></td></tr><tr><td><math>\vdots</math></td><td><math>\vdots</math></td><td><math>\vdots</math></td><td></td><td><math>\vdots</math></td><td><math>\vdots</math></td></tr></table>	$\begin{matrix} & Y \\ X \backslash & \end{matrix}$	$y_1$	$y_2$	$\cdots$	$y_j$	$\cdots$	$x_1$	$p_{11}$	$p_{12}$	$\cdots$	$p_{1j}$	$\cdots$	$x_2$	$p_{21}$	$p_{22}$	$\cdots$	$p_{2j}$	$\cdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$x_i$	$p_{i1}$		$\cdots$	$p_{ij}$	$\cdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
		$\begin{matrix} & Y \\ X \backslash & \end{matrix}$	$y_1$	$y_2$	$\cdots$	$y_j$	$\cdots$																															
		$x_1$	$p_{11}$	$p_{12}$	$\cdots$	$p_{1j}$	$\cdots$																															
		$x_2$	$p_{21}$	$p_{22}$	$\cdots$	$p_{2j}$	$\cdots$																															
		$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$																															
		$x_i$	$p_{i1}$		$\cdots$	$p_{ij}$	$\cdots$																															
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$																																	
这里 $p_{ij}$ 具有下面两个 <b>性质</b> :																																						
(1) $p_{ij} \geq 0$ ( $i,j=1,2,\cdots$ );																																						
(2) $\sum_i \sum_j p_{ij} = 1$ .																																						

	连续型	<p>对于二维随机向量 <math>\xi=(X,Y)</math> , 如果存在非负函数 <math>f(x,y)(-\infty &lt; x &lt; +\infty, -\infty &lt; y &lt; +\infty)</math> , 使对任意一个其邻边分别平行于坐标轴的矩形区域 <math>D</math> , 即 <math>D=\{(X,Y) a &lt; x &lt; b, c &lt; y &lt; d\}</math> 有</p> $P\{(X,Y) \in D\} = \iint_D f(x,y)dx dy,$ <p>则称 <math>\xi</math> 为<b>连续型随机向量</b>; 并称 <math>f(x,y)</math> 为 <math>\xi = (X, Y)</math> 的分布密度或称为 <math>X</math> 和 <math>Y</math> 的<b>联合分布密度</b>。</p> <p>分布密度 <math>f(x,y)</math> 具有下面两个<b>性质</b>:</p> <p>(1) <math>f(x,y) \geq 0</math>;</p> <p>(2) <math>\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)dx dy = 1</math>.</p>
(2) 二维 随机变量 的本质	$\xi(X=x, Y=y) = \xi(X=x \cap Y=y)$	

<p>(3) 联合分布函数</p>	<p>设 <math>(X, Y)</math> 为二维随机变量, 对于任意实数 <math>x, y</math>, 二元函数</p> $F(x, y) = P\{X \leq x, Y \leq y\}$ <p>称为二维随机向量 <math>(X, Y)</math> 的分布函数, 或称为随机变量 <math>X</math> 和 <math>Y</math> 的<b>联合分布函数</b>。</p> <p>分布函数是一个以全平面为其定义域, 以事件 <math>\{(\omega_1, \omega_2)   -\infty &lt; X(\omega_1) \leq x, -\infty &lt; Y(\omega_2) \leq y\}</math> 的概率为函数值的一个实值函数。分布函数 <math>F(x, y)</math> 具有以下的基本<b>性质</b>:</p> <p>(1) <math>0 \leq F(x, y) \leq 1</math>;</p> <p>(2) <math>F(x, y)</math> 分别对 <math>x</math> 和 <math>y</math> 是非减的, 即</p> <p>当 <math>x_2 &gt; x_1</math> 时, 有 <math>F(x_2, y) \geq F(x_1, y)</math>; 当 <math>y_2 &gt; y_1</math> 时, 有 <math>F(x, y_2) \geq F(x, y_1)</math>;</p> <p>(3) <math>F(x, y)</math> 分别对 <math>x</math> 和 <math>y</math> 是右连续的, 即</p> $F(x, y) = F(x + 0, y), F(x, y) = F(x, y + 0);$ <p>(4) <math>F(-\infty, -\infty) = F(-\infty, y) = F(x, -\infty) = 0, F(+\infty, +\infty) = 1</math>.</p> <p>(5) 对于 <math>x_1 &lt; x_2, y_1 &lt; y_2</math>,</p> $F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geq 0.$	
<p>(4) 离散型与连续型的关系</p>	$P(X = x, Y = y) \approx P(x < X \leq x + dx, y < Y \leq y + dy) \approx f(x, y) dx dy$	
<p>(5) 边缘分布</p>	<p>离散型</p>	<p><math>X</math> 的边缘分布为</p> $P_{i\cdot} = P(X = x_i) = \sum_j p_{ij} (i, j = 1, 2, \dots);$ <p><math>Y</math> 的边缘分布为</p> $P_{\cdot j} = P(Y = y_j) = \sum_i p_{ij} (i, j = 1, 2, \dots)。$

	连续型	<p>X 的边缘分布密度为</p> $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy;$ <p>Y 的边缘分布密度为</p> $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$
(6) 条件分布	离散型	<p>在已知 <math>X=x_i</math> 的条件下, Y 取值的条件分布为</p> $P(Y = y_j   X = x_i) = \frac{p_{ij}}{p_{i\bullet}};$ <p>在已知 <math>Y=y_j</math> 的条件下, X 取值的条件分布为</p> $P(X = x_i   Y = y_j) = \frac{p_{ij}}{p_{\bullet j}},$
	连续型	<p>在已知 <math>Y=y</math> 的条件下, X 的条件分布密度为</p> $f(x y) = \frac{f(x, y)}{f_Y(y)};$ <p>在已知 <math>X=x</math> 的条件下, Y 的条件分布密度为</p> $f(y x) = \frac{f(x, y)}{f_X(x)}$
(7) 独立性	一般型	$F(X, Y) = F_X(x)F_Y(y)$
	离散型	$p_{ij} = p_{i\bullet} p_{\bullet j}$ <p>有零不独立</p>
	连续型	$f(x, y) = f_X(x)f_Y(y)$ <p>直接判断, 充要条件:</p> <p>①可分离变量</p> <p>②正概率密度区间为矩形</p>
	二维正态分布	$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]},$ <p><math>\rho = 0</math></p>

	<p>随机变量的函数</p> <p>若 <math>X_1, X_2, \dots, X_m, X_{m+1}, \dots, X_n</math> 相互独立, <math>h, g</math> 为连续函数, 则:</p> <p><math>h(X_1, X_2, \dots, X_m)</math> 和 <math>g(X_{m+1}, \dots, X_n)</math> 相互独立。</p> <p>特例: 若 <math>X</math> 与 <math>Y</math> 独立, 则: <math>h(X)</math> 和 <math>g(Y)</math> 独立。</p> <p>例如: 若 <math>X</math> 与 <math>Y</math> 独立, 则: <math>3X+1</math> 和 <math>5Y-2</math> 独立。</p>
<p>(8) 二维均匀分布</p>	<p>设随机向量 <math>(X, Y)</math> 的分布密度函数为</p> $f(x, y) = \begin{cases} \frac{1}{S_D} & (x, y) \in D \\ 0, & \text{其他} \end{cases}$ <p>其中 <math>S_D</math> 为区域 <math>D</math> 的面积, 则称 <math>(X, Y)</math> 服从 <math>D</math> 上的均匀分布, 记为 <math>(X, Y) \sim U(D)</math>。</p> <p>例如图 3.1、图 3.2 和图 3.3。</p> <div data-bbox="277 936 576 1142"> </div> <p>图 3.1</p> <div data-bbox="277 1227 576 1496"> </div> <p>图 3.2</p> <div data-bbox="277 1603 576 1794"> </div> <p>图 3.3</p>



<p>(9) 二维正态分布</p>	<p>设随机向量 <math>(X, Y)</math> 的分布密度函数为</p> $f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}$ <p>其中 <math>\mu_1, \mu_2, \sigma_1 &gt; 0, \sigma_2 &gt; 0,  \rho  &lt; 1</math> 是 5 个参数, 则称 <math>(X, Y)</math> 服从<b>二维正态分布</b>, 记为 <math>(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)</math>.</p> <p>由边缘密度的计算公式, 可以推出二维正态分布的两个边缘分布仍为正态分布, 即 <math>X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)</math>.</p> <p>但是若 <math>X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)</math>, <math>(X, Y)</math> 未必是二维正态分布。</p>	
<p>(10) 函数分布</p>	<p><math>Z = X + Y</math></p>	<p>根据定义计算: <math>F_Z(z) = P(Z \leq z) = P(X + Y \leq z)</math></p> <p>对于连续型, <math>f_Z(z) = \int_{-\infty}^{+\infty} f(x, z-x)dx</math></p> <p>两个独立的正态分布的和仍为正态分布 <math>(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)</math>。</p> <p><math>n</math> 个相互独立的正态分布的线性组合, 仍服从正态分布。</p> $\mu = \sum_i C_i \mu_i, \quad \sigma^2 = \sum_i C_i^2 \sigma_i^2$
	<p><math>Z = \max\{X_1, X_2, \dots, X_n\}</math></p>	<p>若 <math>X_1, X_2, \dots, X_n</math> 相互独立, 其分布函数分别为 <math>F_{x_1}(x), F_{x_2}(x) \dots F_{x_n}(x)</math>, 则</p> <p><math>Z = \max(X_1, X_2, \dots, X_n)</math> 的分布函数为:</p> $F_{\max}(x) = F_{x_1}(x) \cdot F_{x_2}(x) \cdot \dots \cdot F_{x_n}(x)$ $F_{\min}(x) = 1 - [1 - F_{x_1}(x)] \cdot [1 - F_{x_2}(x)] \cdot \dots \cdot [1 - F_{x_n}(x)]$

	$\chi^2$ 分布	<p>设 <math>n</math> 个随机变量 <math>X_1, X_2, \dots, X_n</math> 相互独立, 且服从标准正态分布, 可以证明它们的平方和</p> $W = \sum_{i=1}^n X_i^2$ <p>的分布密度为</p> $f(u) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} u^{\frac{n}{2}-1} e^{-\frac{u}{2}} & u \geq 0, \\ 0, & u < 0. \end{cases}$ <p>我们称随机变量 <math>W</math> 服从自由度为 <math>n</math> 的 <math>\chi^2</math> 分布, 记为 <math>W \sim \chi^2(n)</math>, 其中</p> $\Gamma\left(\frac{n}{2}\right) = \int_0^{+\infty} x^{\frac{n}{2}-1} e^{-x} dx.$ <p>所谓自由度是指独立正态随机变量的个数, 它是随机变量分布中的一个重要参数。</p> <p><math>\chi^2</math> 分布满足可加性: 设</p> $Y_i \sim \chi^2(n_i),$ <p>则</p> $Z = \sum_{i=1}^k Y_i \sim \chi^2(n_1 + n_2 + \dots + n_k).$
--	-------------	---

t 分布	<p>设 <math>X, Y</math> 是两个相互独立的随机变量, 且</p> $X \sim N(0,1), Y \sim \chi^2(n),$ <p>可以证明函数</p> $T = \frac{X}{\sqrt{Y/n}}$ <p>的概率密度为</p> $f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \quad (-\infty < t < +\infty).$ <p>我们称随机变量 <math>T</math> 服从自由度为 <math>n</math> 的 <math>t</math> 分布, 记为 <math>T \sim t(n)</math>。</p> $t_{1-\alpha}(n) = -t_{\alpha}(n)$
F 分布	<p>设 <math>X \sim \chi^2(n_1), Y \sim \chi^2(n_2)</math>, 且 <math>X</math> 与 <math>Y</math> 独立, 可以证明 <math>F = \frac{X/n_1}{Y/n_2}</math> 的概率密度函数为</p> $f(y) = \begin{cases} \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} y^{\frac{n_1}{2}-1} \left(1 + \frac{n_1}{n_2}y\right)^{-\frac{n_1+n_2}{2}}, & y \geq 0 \\ 0, & y < 0 \end{cases}$ <p>我们称随机变量 <math>F</math> 服从第一个自由度为 <math>n_1</math>, 第二个自由度为 <math>n_2</math> 的 <math>F</math> 分布, 记为 <math>F \sim f(n_1, n_2)</math>。</p> $F_{1-\alpha}(n_1, n_2) = \frac{1}{F_{\alpha}(n_2, n_1)}$

## 第四章 随机变量的数字特征

(1) 一维		离散型	连续型
--------	--	-----	-----

随机变量的数字特征	期望 期望就是平均值	<p>设 <math>X</math> 是离散型随机变量, 其分布律为 <math>P(X = x_k) = p_k</math>, <math>k=1,2,\dots,n</math>,</p> $E(X) = \sum_{k=1}^n x_k p_k$ <p>(要求绝对收敛)</p>	<p>设 <math>X</math> 是连续型随机变量, 其概率密度为 <math>f(x)</math>,</p> $E(X) = \int_{-\infty}^{+\infty} x f(x) dx$ <p>(要求绝对收敛)</p>
	函数的期望	<p><math>Y=g(X)</math></p> $E(Y) = \sum_{k=1}^n g(x_k) p_k$	<p><math>Y=g(X)</math></p> $E(Y) = \int_{-\infty}^{+\infty} g(x) f(x) dx$
	<p>方差</p> <p><math>D(X)=E[X-E(X)]^2</math>,</p> <p>标准差</p> <p><math>\sigma(X) = \sqrt{D(X)}</math>,</p>	$D(X) = \sum_k [x_k - E(X)]^2 p_k$	$D(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x) dx$

	矩	<p>①对于正整数 <math>k</math>, 称随机变量 <math>X</math> 的 <math>k</math> 次幂的数学期望为 <math>X</math> 的 <math>k</math> 阶原点矩, 记为 <math>v_k</math>, 即</p> $v_k = E(X^k) = \sum_i x_i^k p_i, \quad k=1, 2, \dots$ <p>②对于正整数 <math>k</math>, 称随机变量 <math>X</math> 与 <math>E(X)</math> 差的 <math>k</math> 次幂的数学期望为 <math>X</math> 的 <math>k</math> 阶中心矩, 记为 <math>\mu_k</math>, 即</p> $\mu_k = E(X - E(X))^k$ $= \sum_i (x_i - E(X))^k p_i, \quad k=1, 2, \dots$	<p>①对于正整数 <math>k</math>, 称随机变量 <math>X</math> 的 <math>k</math> 次幂的数学期望为 <math>X</math> 的 <math>k</math> 阶原点矩, 记为 <math>v_k</math>, 即</p> $v_k = E(X^k) = \int_{-\infty}^{+\infty} x^k f(x) dx, \quad k=1, 2, \dots$ <p>②对于正整数 <math>k</math>, 称随机变量 <math>X</math> 与 <math>E(X)</math> 差的 <math>k</math> 次幂的数学期望为 <math>X</math> 的 <math>k</math> 阶中心矩, 记为 <math>\mu_k</math>, 即</p> $\mu_k = E(X - E(X))^k$ $= \int_{-\infty}^{+\infty} (x - E(X))^k f(x) dx, \quad k=1, 2, \dots$
	切比雪夫不等式	<p>设随机变量 <math>X</math> 具有数学期望 <math>E(X) = \mu</math>, 方差 <math>D(X) = \sigma^2</math>, 则对于任意正数 <math>\varepsilon</math>, 有下列切比雪夫不等式</p> $P( X - \mu  \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$ <p>切比雪夫不等式给出了在未知 <math>X</math> 的分布的情况下, 对概率</p> $P( X - \mu  \geq \varepsilon)$ <p>的一种估计, 它在理论上具有重要意义。</p>	
(2) 期望的性质	<p>(1) <math>E(C) = C</math></p> <p>(2) <math>E(CX) = CE(X)</math></p> <p>(3) <math>E(X+Y) = E(X) + E(Y)</math>, <math>E(\sum_{i=1}^n C_i X_i) = \sum_{i=1}^n C_i E(X_i)</math></p> <p>(4) <math>E(XY) = E(X)E(Y)</math>, 充分条件: <math>X</math> 和 <math>Y</math> 独立; 充要条件: <math>X</math> 和 <math>Y</math> 不相关。</p>		

(3) 方差的性质	<p>(1) <math>D(C)=0; E(C)=C</math></p> <p>(2) <math>D(aX)=a^2D(X); E(aX)=aE(X)</math></p> <p>(3) <math>D(aX+b)= a^2D(X); E(aX+b)=aE(X)+b</math></p> <p>(4) <math>D(X)=E(X^2)-E^2(X)</math></p> <p>(5) <math>D(X\pm Y)=D(X)+D(Y)</math>, 充分条件: X 和 Y 独立;</p> <p style="text-align: center;">充要条件: X 和 Y 不相关。</p> <p><math>D(X\pm Y)=D(X)+D(Y) \pm 2E[(X-E(X))(Y-E(Y))]</math>, 无条件成立。</p> <p>而 <math>E(X+Y)=E(X)+E(Y)</math>, 无条件成立。</p>		
(4) 常见分布的期望和方差		期望	方差
	0-1 分布 $B(1, p)$	$p$	$p(1-p)$
	二项分布 $B(n, p)$	$np$	$np(1-p)$
	泊松分布 $P(\lambda)$	$\lambda$	$\lambda$
	几何分布 $G(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
	超几何分布 $H(n, M, N)$	$\frac{nM}{N}$	$\frac{nM}{N} \left(1 - \frac{M}{N}\right) \left(\frac{N-n}{N-1}\right)$
	均匀分布 $U(a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
	指数分布 $e(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
	正态分布 $N(\mu, \sigma^2)$	$\mu$	$\sigma^2$
	$\chi^2$ 分布	n	2n
	t 分布	0	$\frac{n}{n-2} (n>2)$

(5) 二维 随机变量 的数字特 征	期望	$E(X) = \sum_{i=1}^n x_i p_{i\bullet}$ $E(Y) = \sum_{j=1}^n y_j p_{\bullet j}$	$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$ $E(Y) = \int_{-\infty}^{+\infty} y f_Y(y) dy$
	函数的期望	$E[G(X, Y)] =$ $\sum_i \sum_j G(x_i, y_j) p_{ij}$	$E[G(X, Y)] =$ $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} G(x, y) f(x, y) dx dy$
	方差	$D(X) = \sum_i [x_i - E(X)]^2 p_{i\bullet}$ $D(Y) = \sum_j [y_j - E(Y)]^2 p_{\bullet j}$	$D(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f_X(x) dx$ $D(Y) = \int_{-\infty}^{+\infty} [y - E(Y)]^2 f_Y(y) dy$
	协方差	<p>对于随机变量 X 与 Y, 称它们的二阶混合中心矩 <math>\mu_{11}</math> 为 X 与 Y 的协方差或相关矩, 记为 <math>\sigma_{XY}</math> 或 <math>\text{cov}(X, Y)</math>, 即</p> $\sigma_{XY} = \mu_{11} = E[(X - E(X))(Y - E(Y))].$ <p>与记号 <math>\sigma_{XY}</math> 相对应, X 与 Y 的方差 D (X) 与 D (Y) 也可分别记为 <math>\sigma_{XX}</math> 与 <math>\sigma_{YY}</math>。</p>	

	相关系数	<p>对于随机变量 <math>X</math> 与 <math>Y</math>, 如果 <math>D(X) &gt; 0, D(Y) &gt; 0</math>, 则称</p> $\frac{\sigma_{XY}}{\sqrt{D(X)}\sqrt{D(Y)}}$ <p>为 <math>X</math> 与 <math>Y</math> 的相关系数, 记作 <math>\rho_{XY}</math> (有时可简记为 <math>\rho</math> )。</p> <p><math> \rho  \leq 1</math>, 当 <math> \rho  = 1</math> 时, 称 <math>X</math> 与 <math>Y</math> 完全相关: <math>P(X = aY + b) = 1</math></p> <p>完全相关 <math>\begin{cases} \text{正相关, 当 } \rho = 1 \text{ 时 } (a &gt; 0), \\ \text{负相关, 当 } \rho = -1 \text{ 时 } (a &lt; 0), \end{cases}</math></p> <p>而当 <math>\rho = 0</math> 时, 称 <math>X</math> 与 <math>Y</math> 不相关。</p> <p>以下五个命题是等价的:</p> <p>① <math>\rho_{XY} = 0</math>;</p> <p>② <math>\text{cov}(X, Y) = 0</math>;</p> <p>③ <math>E(XY) = E(X)E(Y)</math>;</p> <p>④ <math>D(X+Y) = D(X) + D(Y)</math>;</p> <p>⑤ <math>D(X-Y) = D(X) + D(Y)</math>.</p>
	协方差矩阵	$\begin{pmatrix} \sigma_{XX} & \sigma_{XY} \\ \sigma_{YX} & \sigma_{YY} \end{pmatrix}$
	混合矩	<p>对于随机变量 <math>X</math> 与 <math>Y</math>, 如果有 <math>E(X^k Y^l)</math> 存在, 则称之为 <math>X</math> 与 <math>Y</math> 的 <math>k+l</math> 阶混合原点矩, 记为 <math>\nu_{kl}</math>; <math>k+l</math> 阶混合中心矩记为:</p> $u_{kl} = E[(X - E(X))^k (Y - E(Y))^l].$
(6) 协方差的性质	<p>(1) <math>\text{cov}(X, Y) = \text{cov}(Y, X)</math>;</p> <p>(2) <math>\text{cov}(aX, bY) = ab \text{cov}(X, Y)</math>;</p> <p>(3) <math>\text{cov}(X_1 + X_2, Y) = \text{cov}(X_1, Y) + \text{cov}(X_2, Y)</math>;</p> <p>(4) <math>\text{cov}(X, Y) = E(XY) - E(X)E(Y)</math>.</p>	



(7) 独立 和不相关	<p>(1) 若随机变量 <math>X</math> 与 <math>Y</math> 相互独立, 则 <math>\rho_{XY} = 0</math>; 反之不真。</p> <p>(2) 若 <math>(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)</math>, 则 <math>X</math> 与 <math>Y</math> 相互独立的充要条件是 <math>X</math> 和 <math>Y</math> 不相关。</p>
----------------	---

## 第五章 大数定律和中心极限定理

(1) 大数定律 $\bar{X} \rightarrow \mu$	切比雪夫 大数定律	<p>设随机变量 <math>X_1, X_2, \dots</math> 相互独立, 均具有有限方差, 且被同一常数 <math>C</math> 所界: <math>D(X_i) &lt; C (i=1, 2, \dots)</math>, 则对于任意的正数 <math>\varepsilon</math>, 有</p> $\lim_{n \rightarrow \infty} P\left(\left \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right  < \varepsilon\right) = 1.$ <p>特殊情形: 若 <math>X_1, X_2, \dots</math> 具有相同的数学期望 <math>E(X_i) = \mu</math>, 则上式成为</p> $\lim_{n \rightarrow \infty} P\left(\left \frac{1}{n} \sum_{i=1}^n X_i - \mu\right  < \varepsilon\right) = 1.$
	伯努利大 数定律	<p>设 <math>\mu</math> 是 <math>n</math> 次独立试验中事件 <math>A</math> 发生的次数, <math>p</math> 是事件 <math>A</math> 在每次试验中发生的概率, 则对于任意的正数 <math>\varepsilon</math>, 有</p> $\lim_{n \rightarrow \infty} P\left(\left \frac{\mu}{n} - p\right  < \varepsilon\right) = 1.$ <p>伯努利大数定律说明, 当试验次数 <math>n</math> 很大时, 事件 <math>A</math> 发生的频率与概率有较大判别的可能性很小, 即</p> $\lim_{n \rightarrow \infty} P\left(\left \frac{\mu}{n} - p\right  \geq \varepsilon\right) = 0.$ <p>这就以严格的数学形式描述了频率的稳定性。</p>
	辛钦大数 定律	<p>设 <math>X_1, X_2, \dots, X_n, \dots</math> 是相互独立同分布的随机变量序列, 且 <math>E(X_n) = \mu</math>, 则对于任意的正数 <math>\varepsilon</math> 有</p> $\lim_{n \rightarrow \infty} P\left(\left \frac{1}{n} \sum_{i=1}^n X_i - \mu\right  < \varepsilon\right) = 1.$

<p>(2) 中心极限定理</p> $\bar{X} \rightarrow N(\mu, \frac{\sigma^2}{n})$	<p>列维 - 林德伯格定理</p>	<p>设随机变量 <math>X_1, X_2, \dots</math> 相互独立, 服从同一分布, 且具有相同的数学期望和方差: <math>E(X_k) = \mu, D(X_k) = \sigma^2 \neq 0 (k = 1, 2, \dots)</math>, 则随机变量</p> $Y_n = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma}$ <p>的分布函数 <math>F_n(x)</math> 对任意的实数 <math>x</math>, 有</p> $\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} P\left\{ \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \leq x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$ <p>此定理也称为独立同分布的中心极限定理。</p>
	<p>棣莫弗 - 拉普拉斯定理</p>	<p>设随机变量 <math>X_n</math> 为具有参数 <math>n, p (0 &lt; p &lt; 1)</math> 的二项分布, 则对于任意实数 <math>x</math>, 有</p> $= \lim_{n \rightarrow \infty} P\left\{ \frac{X_n - np}{\sqrt{np(1-p)}} \leq x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$
<p>(3) 二项定理</p>	<p>若当 <math>N \rightarrow \infty</math> 时, <math>\frac{M}{N} \rightarrow p (n, k \text{ 不变})</math>, 则</p> $\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \rightarrow C_n^k p^k (1-p)^{n-k} \quad (N \rightarrow \infty).$ <p>超几何分布的极限分布为二项分布。</p>	
<p>(4) 泊松定理</p>	<p>若当 <math>n \rightarrow \infty</math> 时, <math>np \rightarrow \lambda &gt; 0</math>, 则</p> $C_n^k p^k (1-p)^{n-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda} \quad (n \rightarrow \infty).$ <p>其中 <math>k=0, 1, 2, \dots, n, \dots</math>。</p> <p>二项分布的极限分布为泊松分布。</p>	

## 第六章 样本及抽样分布

<p>(1) 数理统计的基</p>	<p>总体</p>	<p>在数理统计中, 常把被考察对象的某一个 (或多个) 指标的全体称为 <b>总体</b> (或母体)。我们总是把总体看成一个具有分布的<b>随机变量</b> (或<b>随机向量</b>)。</p>
-------------------	-----------	--

本概念	个体	总体中的每一个单元称为 <b>样品</b> （或 <b>个体</b> ）。
	样本	我们把从总体中抽取的部分样品 $x_1, x_2, \dots, x_n$ 称为 <b>样本</b> 。样本中所含的样品数称为样本容量，一般用 $n$ 表示。在一般情况下，总是把样本看成是 $n$ 个相互独立的且与总体有相同分布的随机变量，这样的样本称为简单随机样本。在泛指任一次抽取的结果时， $x_1, x_2, \dots, x_n$ 表示 $n$ 个随机变量（样本）；在具体的一次抽取之后， $x_1, x_2, \dots, x_n$ 表示 $n$ 个具体的数值（样本值）。我们称之为样本的两重性。
	样本函数和统计量	<p>设 <math>x_1, x_2, \dots, x_n</math> 为总体的一个样本，称</p> $\varphi = \varphi(x_1, x_2, \dots, x_n)$ <p>为<b>样本函数</b>，其中 <math>\varphi</math> 为一个连续函数。如果 <math>\varphi</math> 中不包含任何未知参数，则称 <math>\varphi(x_1, x_2, \dots, x_n)</math> 为一个<b>统计量</b>。</p>

	常见统计量及其性质	<p>样本均值 <math>\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.</math></p> <p>样本方差 <math>S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.</math></p> <p>样本标准差 <math>S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.</math></p> <p>样本 k 阶<b>原点矩</b></p> <p><math>M_k = \frac{1}{n} \sum_{i=1}^n x_i^k, k=1,2,\dots.</math></p> <p>样本 k 阶<b>中心矩</b></p> <p><math>M'_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k, k=2,3,\dots.</math></p> <p><math>E(\bar{X}) = \mu, D(\bar{X}) = \frac{\sigma^2}{n},</math></p> <p><math>E(S^2) = \sigma^2, E(S^{*2}) = \frac{n-1}{n} \sigma^2,</math></p> <p>其中 <math>S^{*2} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2</math>, 为<b>二阶中心矩</b>。</p>
(2) 正态总体下的四大分布	正态分布	<p>设 <math>x_1, x_2, \dots, x_n</math> 为来自正态总体 <math>N(\mu, \sigma^2)</math> 的一个样本, 则样本函数</p> <p><math>u \stackrel{\text{def}}{=} \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1).</math></p>
	t 分布	<p>设 <math>x_1, x_2, \dots, x_n</math> 为来自正态总体 <math>N(\mu, \sigma^2)</math> 的一个样本, 则样本函数</p> <p><math>t \stackrel{\text{def}}{=} \frac{\bar{x} - \mu}{s / \sqrt{n}} \sim t(n-1),</math></p> <p>其中 <math>t(n-1)</math> 表示自由度为 <math>n-1</math> 的 t 分布。</p>
	$\chi^2$ 分布	<p>设 <math>x_1, x_2, \dots, x_n</math> 为来自正态总体 <math>N(\mu, \sigma^2)</math> 的一个样本, 则样本函数</p> <p><math>w \stackrel{\text{def}}{=} \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),</math></p> <p>其中 <math>\chi^2(n-1)</math> 表示自由度为 <math>n-1</math> 的 <math>\chi^2</math> 分布。</p>

	F 分布	<p>设 <math>x_1, x_2, \dots, x_n</math> 为来自正态总体 <math>N(\mu, \sigma_1^2)</math> 的一个样本, 而 <math>y_1, y_2, \dots, y_n</math> 为来自正态总体 <math>N(\mu, \sigma_2^2)</math> 的一个样本, 则样本函数</p> $F \stackrel{\text{def}}{=} \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1),$ <p>其中</p> $S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2, \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2;$ <p><math>F(n_1 - 1, n_2 - 1)</math> 表示第一自由度为 <math>n_1 - 1</math>, 第二自由度为 <math>n_2 - 1</math> 的 F 分布。</p>
(3) 正态 总体下分 布的性质	$\bar{X}$ 与 $S^2$ 独立。	

## 第七章 参数估计



	极大似然估计	<p>当总体 <math>X</math> 为连续型随机变量时, 设其分布密度为 <math>f(x; \theta_1, \theta_2, \dots, \theta_m)</math>, 其中 <math>\theta_1, \theta_2, \dots, \theta_m</math> 为未知参数。又设 <math>x_1, x_2, \dots, x_n</math> 为总体的一个样本, 称</p> $L(\theta_1, \theta_2, \dots, \theta_m) = \prod_{i=1}^n f(x_i; \theta_1, \theta_2, \dots, \theta_m)$ <p>为样本的<b>似然函数</b>, 简记为 <math>L_n</math>。</p> <p>当总体 <math>X</math> 为离型随机变量时, 设其分布律为 <math>P\{X = x\} = p(x; \theta_1, \theta_2, \dots, \theta_m)</math>, 则称</p> $L(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_m) = \prod_{i=1}^n p(x_i; \theta_1, \theta_2, \dots, \theta_m)$ <p>为样本的<b>似然函数</b>。</p> <p>若似然函数 <math>L(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_m)</math> 在 <math>\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m</math> 处取到最大值, 则称 <math>\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m</math> 分别为 <math>\theta_1, \theta_2, \dots, \theta_m</math> 的最大似然估计值, 相应的统计量称为<b>最大似然估计量</b>。</p> $\left. \frac{\partial \ln L_n}{\partial \theta_i} \right _{\theta_i = \hat{\theta}_i} = 0, i = 1, 2, \dots, m$ <p>若 <math>\hat{\theta}</math> 为 <math>\theta</math> 的极大似然估计, <math>g(x)</math> 为单调函数, 则 <math>g(\hat{\theta})</math> 为 <math>g(\theta)</math> 的<b>极大似然估计</b>。</p>
(2) 估 计 量 的 评 选 标准	无偏性	<p>设 <math>\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)</math> 为未知参数 <math>\theta</math> 的估计量。若 <math>E(\hat{\theta}) = \theta</math>, 则称 <math>\hat{\theta}</math> 为 <math>\theta</math> 的<b>无偏估计量</b>。</p> $E(\bar{X}) = E(X), \quad E(S^2) = D(X)$
	有效性	<p>设 <math>\hat{\theta}_1 = \hat{\theta}_1(x_1, x_2, \dots, x_n)</math> 和 <math>\hat{\theta}_2 = \hat{\theta}_2(x_1, x_2, \dots, x_n)</math> 是未知参数 <math>\theta</math> 的两个无偏估计量。若 <math>D(\hat{\theta}_1) &lt; D(\hat{\theta}_2)</math>, 则称 <math>\hat{\theta}_1</math> <b>比</b> <math>\hat{\theta}_2</math> <b>有效</b>。</p>

	一致性	<p>设<math>\hat{\theta}_n</math>是<math>\theta</math>的一串估计量, 如果对于任意的正数<math>\varepsilon</math>, 都有</p> $\lim_{n \rightarrow \infty} P( \hat{\theta}_n - \theta  > \varepsilon) = 0,$ <p>则称<math>\hat{\theta}_n</math>为<math>\theta</math>的一致估计量 (或相合估计量)。</p> <p>若<math>\hat{\theta}</math> 为<math>\theta</math>的无偏估计, 且<math>D(\hat{\theta}) \rightarrow 0 (n \rightarrow \infty)</math>, 则<math>\hat{\theta}</math>为<math>\theta</math>的一致估计。</p> <p>只要总体的 <math>E(X)</math>和 <math>D(X)</math>存在, 一切样本矩和样本矩的连续函数都是相应总体的一致估计量。</p>	
(3) 区 间 估 计	置信区 间和置 信度	<p>设总体 <math>X</math> 含有一个待估的未知参数<math>\theta</math>。如果我们从样本<math>x_1, x_2, \dots, x_n</math>出发, 找出两个统计量<math>\theta_1 = \theta_1(x_1, x_2, \dots, x_n)</math>与<math>\theta_2 = \theta_2(x_1, x_2, \dots, x_n) (\theta_1 &lt; \theta_2)</math>, 使得区间<math>[\theta_1, \theta_2]</math>以<math>1 - \alpha (0 &lt; \alpha &lt; 1)</math>的概率包含这个待估参数<math>\theta</math>, 即</p> $P\{\theta_1 \leq \theta \leq \theta_2\} = 1 - \alpha,$ <p>那么称区间<math>[\theta_1, \theta_2]</math>为<math>\theta</math>的置信区间, <math>1 - \alpha</math>为该区间的置信度 (或置信水平)。</p>	
	单正态 总体的 期望和 方差的 区间估 计	<p>设<math>x_1, x_2, \dots, x_n</math>为总体<math>X \sim N(\mu, \sigma^2)</math>的一个样本, 在置信度为<math>1 - \alpha</math>下, 我们来确定<math>\mu</math>和<math>\sigma^2</math>的置信区间<math>[\theta_1, \theta_2]</math>。具体步骤如下:</p> <p>(i) 选择样本函数;</p> <p>(ii) 由置信度<math>1 - \alpha</math>, 查表找分位数;</p> <p>(iii) 导出置信区间<math>[\theta_1, \theta_2]</math>。</p>	
		已知方差, 估计均值	<p>(i) 选择样本函数</p> $u = \frac{\bar{x} - \mu}{\sigma_0/\sqrt{n}} \sim N(0,1).$ <p>(ii) 查表找分位数</p> $P\left(-\lambda \leq \frac{\bar{x} - \mu}{\sigma_0/\sqrt{n}} \leq \lambda\right) = 1 - \alpha.$ <p>(iii) 导出置信区间</p> $\left[\bar{x} - \lambda \frac{\sigma_0}{\sqrt{n}}, \bar{x} + \lambda \frac{\sigma_0}{\sqrt{n}}\right]$



		未知方差，估计均值	<p>(i) 选择样本函数</p> $t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim t(n-1).$ <p>(ii) 查表找分位数</p> $P\left(-\lambda \leq \frac{\bar{x} - \mu}{S/\sqrt{n}} \leq \lambda\right) = 1 - \alpha.$ <p>(iii) 导出置信区间</p> $\left[\bar{x} - \lambda \frac{S}{\sqrt{n}}, \bar{x} + \lambda \frac{S}{\sqrt{n}}\right]$
		方差的区间估计	<p>(i) 选择样本函数</p> $w = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$ <p>(ii) 查表找分位数</p> $P\left(\lambda_1 \leq \frac{(n-1)S^2}{\sigma^2} \leq \lambda_2\right) = 1 - \alpha.$ <p>(iii) 导出<math>\sigma</math>的置信区间</p> $\left[\sqrt{\frac{n-1}{\lambda_2}} S, \sqrt{\frac{n-1}{\lambda_1}} S\right]$

## 第八章 假设检验

基 本 思 想	<p>假设检验的统计思想是，概率很小的事件在一次试验中可以认为基本上是不会发生的，即小概率原理。</p> <p>为了检验一个假设 <math>H_0</math> 是否成立。我们先假定 <math>H_0</math> 是成立的。如果根据这个假定导致了一个不合理的事件发生，那就表明原来的假定 <math>H_0</math> 是不正确的，我们拒绝接受 <math>H_0</math>；如果由此没有导出不合理的现象，则不能拒绝接受 <math>H_0</math>，我们称 <math>H_0</math> 是相容的。与 <math>H_0</math> 相对的假设称为备择假设，用 <math>H_1</math> 表示。</p> <p>这里所说的小概率事件就是事件 <math>\{K \in R_\alpha\}</math>，其概率就是检验水平 <math>\alpha</math>，通常我们取 <math>\alpha=0.05</math>，有时也取 0.01 或 0.10。</p>
------------	--

基 本 步 骤	<p>假设检验的基本步骤如下：</p> <p>(1) 提出零假设 <math>H_0</math>;</p> <p>(2) 选择统计量 <math>K</math>;</p> <p>(3) 对于检验水平<math>\alpha</math>查表找分位数<math>\lambda</math>;</p> <p>(4) 由样本值<math>x_1, x_2, \dots, x_n</math>计算统计量之值 <math>K</math>;</p> <p>将<math>\hat{K}</math>与<math>\lambda</math>进行比较, 作出判断: 当<math> \hat{K}  &gt; \lambda</math>(或<math>\hat{K} &gt; \lambda</math>)时否定 <math>H_0</math>, 否则认为 <math>H_0</math> 相容。</p>	
两 类 错 误	第一类错误	<p>当 <math>H_0</math> 为真时, 而样本值却落入了否定域, 按照我们规定的检验法则, 应当否定 <math>H_0</math>。这时, 我们把客观上 <math>H_0</math> 成立判为 <math>H_0</math> 为不成立 (即否定了真实的假设), 称这种错误为“以真当假”的错误或第一类错误, 记 <math>\alpha</math> 为犯此类错误的概率, 即</p> $P\{\text{否定 } H_0   H_0 \text{ 为真}\} = \alpha;$ <p>此处的<math>\alpha</math>恰好为检验水平。</p>
	第二类错误	<p>当 <math>H_1</math> 为真时, 而样本值却落入了相容域, 按照我们规定的检验法则, 应当接受 <math>H_0</math>。这时, 我们把客观上 <math>H_0</math>。不成立判为 <math>H_0</math> 成立 (即接受了不真实的假设), 称这种错误为“以假当真”的错误或第二类错误, 记<math>\beta</math>为犯此类错误的概率, 即</p> $P\{\text{接受 } H_0   H_1 \text{ 为真}\} = \beta。$

	两类错误的关系	<p>人们当然希望犯两类错误的概率同时都很小。但是，当容量 <math>n</math> 一定时，<math>\alpha</math> 变小，则 <math>\beta</math> 变大；相反地，<math>\beta</math> 变小，则 <math>\alpha</math> 变大。取定 <math>\alpha</math> 要想使 <math>\beta</math> 变小，则必须增加样本容量。</p> <p>在实际使用时，通常人们只能控制犯第一类错误的概率，即给定显著性水平 <math>\alpha</math>。<math>\alpha</math> 大小的选取应根据实际情况而定。当我们宁可“以假为真”、而不愿“以真当假”时，则应把 <math>\alpha</math> 取得很小，如 0.01，甚至 0.001。反之，则应把 <math>\alpha</math> 取得大些。</p>
--	---------	---

### 单正态总体均值和方差的假设检验

条件	零假设	统计量	对应样本函数分布	否定域
已知 $\sigma^2$	$H_0: \mu = \mu_0$	$U = \frac{\bar{x} - \mu_0}{\sigma_0/\sqrt{n}}$	$N(0, 1)$	$ u  > u_{1-\frac{\alpha}{2}}$
	$H_0: \mu \leq \mu_0$			$u > u_{1-\alpha}$
	$H_0: \mu \geq \mu_0$			$u < -u_{1-\alpha}$
未知 $\sigma^2$	$H_0: \mu = \mu_0$	$T = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$	$t(n-1)$	$ t  > t_{1-\frac{\alpha}{2}}(n-1)$
	$H_0: \mu \leq \mu_0$			$t > t_{1-\alpha}(n-1)$
	$H_0: \mu \geq \mu_0$			$t < -t_{1-\alpha}(n-1)$
未知 $\sigma^2$	$H_0: \sigma^2 = \sigma_0^2$	$w = \frac{(n-1)S^2}{\sigma_0^2}$	$\kappa^2(n-1)$	$w < \kappa_{\frac{\alpha}{2}}^2(n-1)$ 或 $w > \kappa_{1-\frac{\alpha}{2}}^2(n-1)$
	$H_0: \sigma^2 \leq \sigma_0^2$			$w > \kappa_{1-\alpha}^2(n-1)$
	$H_0: \sigma^2 \geq \sigma_0^2$			$w < \kappa_{\alpha}^2(n-1)$