Lecture 1: Random Experiment, Sample Space and Event

1. 自然界所观察到的现象：**确定性现象**和**随机现象**

- 在一定条件下必然出现的现象，称为**确定性现象**；
- 在一定条件下可能出现也可能不出现的现象，称为**随机现象**。

概率论是研究随机现象规律性的一门数学学科。

2. 定义：在概率论中, 把具有以下三个特征的试验称为随机试验（Random Experiment）。

(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。

3. 定义：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**（Sample Space），记为 S。样本空间的元素，即试验 E 的每一个结果，称为**样本点**（Sample Point）。

4. 定义：随机试验 E 的样本空间 S 的子集称为 E 的**随机事件** (Random Event), 简称**事件** (Event).

5. 随机事件的分类：

- 只含一个样本点的事件称为**基本事件**（Elementary Event）
- 含有多于一个样本点的事件称为**复合事件**（Compound Event）
- Ω：**必然事件**（Certain Event）
- $\emptyset$：**不可能事件** (Impossible Event)

必然事件的对立面是不可能事件，不可能事件的对立面是必然事件，它们互称为**对立事件**（Complementary Event）。

6. 事件的关系：设试验 E 的样本空间为 S，而 A，B，$A_k(k=1,2,...)$ 是 S 的子集。

(1) **包含关系**：若事件 A 出现，必然导致 B 出现，则称事件 B 包含事件 A，记作 $A \subset B$ (或$B \supset A$)；

(2) **A 等于 B**：若事件 A 包含事件 B，而且事件 B 包含事件 A，则称事件 A 与事件 B 相等，记作 A=B.

(3) **事件 A 与 B 的并 (和事件)**：”事件 A、B 至少有一个发生”称为事件 A 与 B 的**和**或**并** (union)，记作 $A \cup B$ (或$A+B$)；
推广：如 n 个事件的并

$$\bigcup_{i=1}^{n} A_i = \text{“事件 } A_1, \cdots, A_n \text{ 至少有一个发生”}$$

可数个事件的并

$$\bigcup_{i=1}^{\infty} A_i = \text{“事件 } A_1, A_2, \cdots \text{ 至少有一个发生”}$$

(4) **事件 A 与 B 的交 (积事件)**：“事件 A、B 同时发生”称为事件 A 与 B 的**积**或**交** (intersection)，记作 AB (或$A \cap B$)；推广：n 个事件的交

$$\bigcap_{i=1}^{n} A_i = \text{“事件 } A_1, \cdots, A_n \text{ 全都发生”}$$

可数个事件的交

$$\bigcap_{i=1}^{\infty} A_i = \text{“事件 } A_1, A_2, \cdots \text{ 全都发生”}$$

(5) **事件 A 与 B 互不相容 (互斥)**：若事件 A 的出现必然导致事件 B 不出现，B 出现也必然导致 A 不出现, 则称事件 A 与 B 互不相容，即 $A \cap B = AB = \emptyset$;

(6) **事件 A 与 B 的差**：由事件 A 出现而事件 B 不出现所组成的事件称为事件 A 与 B 的差，记作 $A - B$;

(7) **事件 A 的对立事件**：设 A 表示“事件 A 出现”，则“事件 A 不出现”称为事件 A 的对立事件或逆事件，记作 $\overline{A}$。若 A 与 B 互逆, 则有 $A \cup B = S$ $AB = \emptyset$。

7. 事件的运算：设 A，B，C 为事件，则有

(1) 交换律：$AB = BA$，$A \cup B = B \cup A$

(2) 结合律：$(AB)C = A(BC)$，$(A \cup B) \cup C = A \cup (B \cup C)$

(3) 分配律：$A(B \cup C) = AB \cup AC$，$A(B - C) = AB - AC$

(4) 德 · 摩根律（对偶律)：$\overline{A \cap B} = \overline{A} \cup \overline{B}$，$\overline{A \cup B} = \overline{A} \cap \overline{B}$

Lecture 2: Frequency and Probability

1. 在相同条件下，进行了 n 次试验，在这 n 次试验中，事件 A 发生的次数 n_A 被成为事件 A 发生的**频数**，比值 $\frac{n_A}{n}$ 被称为事件 A 发生的**频率** (Frequency)，并计作 $f_n(A)$.

2. 在相同条件下重复进行的试验中，若随着试验次数 n 的增加，事件 A 发生的频率稳定在某一常数 p 附近，则称 p 为事件 A 的**概率**，记作 $P(A) = p$。

3. 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 $P(A)$。若集合函数 $P(\cdot)$ 满足下列条件：

(1) 非负性 (Nonnegativity)：对于每一个事件 A，有 $P(A) \geq 0$；

(2) 规范性 (Normalization)：对于必然事件 S，有 $P(S) = 1$；

(3) 可列可加性 (Countable Additivity)：设 $A_1, A_2, \cdots$ 是两两互不相容的事件，则

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

则称 $P(A)$ 为事件 A 的概率 (Probability)。

4. 概率的主要性质：

(1) $P(\emptyset) = 0$；

(2) 若 $A_1, A_2, \cdots, A_n$ 是两两互不相容的事件，则 $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$；

(3) 设 A，B 为两个事件，且 $A \subset B$，则 $P(A) < P(B), P(B-A) = P(B) - P(A)$；

(4) 对于任一事件 A，$P(A) < 1$；

(5) 设 $\overline{A}$ 是 A 的对立事件，则 $P(\overline{A}) = 1 - P(A)$；

(6) (加法公式)对于任意两事件 A,B 有 $P(A \cup B) = P(A) + P(B) - P(AB)$。

Lecture 3: Counting Rule

1. **等可能概率模型（古典概型）**：如果一个随机试验 E 具有以下特点：（1）试验的样本空间只包含有限个样本点；（2）试验中每个基本事件/样本点发生的可能性相等。则称此随机试验是古典型的．此时对每个事件 $A \subset \Omega$，

$$P(A) = \frac{\text{事件 } A \text{ 包含样本点的个数}}{\text{样本点总数}} = \frac{n(A)}{n(\Omega)}$$

称为事件 A 的**古典概率**。

2. **几何概型**：当随机试验的样本空间是某个区域，并且任意一点落在度量 (长度、面积、体积) 相同的子区域是等可能的，则事件 A 的概率可定义为

$$P(A) = \frac{S_A}{S}$$

。

Lecture 4: Conditional Probability

1. 设 A、B 是两个事件，且 $P(A) > 0$，称

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

为事件 A 发生的条件下事件 B 发生的**条件概率**（Conditional Probability）。

2. 条件概率的主要性质：

(1) 非负性：$P(B|A) \geq 0$;

(2) 规范性：$P(S|B) = 1, P(\emptyset|B) = 0$;

(3) $P(B_1 \cup B_2) = P(B_1|A) + P(B_2|A) - P(B_1B_2|A)$;

(4) $P(B|A) = 1 - P(\overline{A}|B)$;

(5) 可列可加性：设 $B_1, B_2, \cdots$ 是两两互不相容的事件，则有

$$P\left(\bigcup_{i=1}^{\infty} B_i|A\right) = \sum_{i=1}^{\infty} P(B_i|A)$$

。

3. **乘法原理**（Multiplication Rule）：由条件概率的定义，得到

- 若 $P(A) > 0$，则有 $P(AB) = P(A)P(B|A)$;
- 若 $P(AB) > 0$，则有 $P(ABC) = P(C|AB)P(B|A)P(A)$;
- 若 $P(A_1A_2 \cdots A_{n-1}) > 0$，则有

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

4. 定义：设 S 为试验 E 的样本空间，$A_1, A_2, \cdots, A_n$ 为 E 的一组事件。若以下条件成立：

1. $A_1, A_2, \cdots, A_n$ 两两互斥：$A_i, A_j = \emptyset, i \neq j, i, j = 1, 2, \cdot, n$;
2. $A_1 \cup A_2 \cup \cdots \cup A_n = S$,

则称 $A_1, A_2, \cdots, A_n$ 为样本空间 S 的一个**划分**（Partition）。

5. **全概率公式**（Law of Total Probability）：设试验 E 的样本空间为 S，B 为 E 的事件，$A_1, A_2, \cdots, A_n$ 为 S 的一个划分，且 $P(A_i) > 0 (i = 1, 2, ..., n)$，则

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i)$$

6. **贝叶斯公式**（Bayes' Theorem）：设试验 E 的样本空间为 S，B 为 E 的事件，$A_1, A_2, \cdots, A_n$ 为 S 的一个划分，且 $P(B) > 0, P(A_i) > 0 (i = 1, 2, ..., n)$，则

$$P(A_i|B) = \frac{P(B|A_i) P(A_i)}{\sum_{i=1}^{n} P(B|A_i) P(A_i)}$$

Lecture 5: Independence

1. 若两事件 A、B 满足 $P(AB) = P(A)P(B)$，则称**事件 A、B 相互独立**（Independence）。

2. 若三个事件 A、B、C 满足 $P(AB|C) = P(A|C)P(B|C), P(C) > 0$，则称事件 A、B 条件独立（Conditional Independence）。

3. 设 A, B, C 是三个事件，如果满足等式

$$\begin{cases} P(AB) = P(A)P(B) \\ P(AC) = P(A)P(C) \\ P(BC) = P(B)P(C) \end{cases}$$

则称事件 A, B, C **两两相互独立**（Pairwise Independence）。

4. 设 A, B, C 是三个事件，如果满足等式

$$\begin{cases} P(AB) = P(A)P(B) \\ P(AC) = P(A)P(C) \\ P(BC) = P(B)P(C) \\ P(ABC) = P(A)P(B)P(C) \end{cases}$$

则称事件 A, B, C **相互独立**。

5. 设 $A_1, A_2, \cdots, A_n$ 是 n 个事件，如果对于任意 $k(1 < k \leq n)$，任意 $1 \leq i_1 < i_2 < \cdot < i_k \leq n$，具有等式

$$P(A_{i_1} A_{i_2}, \cdot, A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdot P(A_{i_k}).$$

则称 $A_1, A_2, \cdots, A_n$ 为相互独立的事件。

6. 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。

7. 设 $n(n \geq 2)$ 个事件 $A_1, A_2, \cdots, A_n$ 相互独立，则

(1) 其中任意 $k(k \geq 2)$ 个事件也是相互独立的；

(2) 将若干个 A_i 用 $\overline{A}_i$ 替换后，得到的新事件集也相互独立；

(3) 特别地，我们有

$$P(A_1 \cup \cdots \cup A_n) = 1 - P(\overline{A}_1) \cdots P(\overline{A}_n).$$

Lecture 6: Discrete Random Variable

1. 设 E 是随机试验，它的样本空间是 $S = e$。如果对于每一个 $e \in S$，又一个实属 $X(e)$ 与之对应，这样就得到一个定义在 S 上的单值实值函数 $X(e)$，称 $X(e)$ 为**随机变量**（Random Variable, RV）。

2. 设离散型随机变量 X 所有可能取的值为 $x_k(k = 1, 2, \cdots)$，X 各个可能值的概率，即事件 $X = x_k$ 的概率，为 $P(X = x_k) = p_k$，$k = 1, 2, \cdots$. 称此为离散型随机变量 X 的分布律。称离散型随机变量特定值的取值为概率质量函数（Probability Mass Function, PMF）：$p_X(x) = P(X = x)$。

3. 分布律 $\{p_k\}$ 的主要性质：$\begin{cases} p_k \geq 0, \quad k = 1, 2, \cdots \\ \sum\limits_{k}^{\infty} p_k = 1 \end{cases}$

4. 常见的离散型分布：

(1) 若随机变量 X 只可能取 0 和 1 两个值，它的分布律为

X	0	1
P	$1-p$	p

，则称 X 服从参数为 p 的**两点分布**（或 **0-1 分布**）.

(2) 若随机变量 X 的分布律为

X	a_1	a_2	$\cdots$	a_n
P_k	$1/n$	$1/n$	$\cdots$	$1/n$

，其中 $a_i \neq a_j, (i \neq j)$，则称 X 服从等可能分布。

(3) 若随机变量 X 的分布律为

$$P\{X = k\} = p(1-p)^{k-1}, \quad k = 1, 2, 3 \cdots,$$

则称 X 服从参数为 p 的**几何分布**（Geometric Distribution），记为 $X \sim G(p)$.

(4) 如果随机变量 X 服从以下分布律

$$P\{X = k\} = b(k; n, p) = C_n^k p^k (1-p)^{n-k},$$

其中 $0 < p < 1$，$0 \leq k \leq n$，则称 X 服从参数为 n，p 的**二项分布**（Binomial Distribution），记为 $\boldsymbol{X} \sim B(n, p)$.

(5) 如果随机变量 X 服从以下分布律

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

其中 $\lambda>0$, 则称 X 服从参数为 λ 的**泊松分布** (Poisson Distribution), 记为 $\boldsymbol{X}\sim P(\lambda)$.

Lecture 7: Continuous Random Variable

1. 设 X 是一个随机变量，x 是任意实数，函数 $F(x) = P\{X \leq x\}$ 成为 X 的**分布函数**（Cumulative Distribution Function，CDF）。

2. 分布函数的主要性质：

(1) $0 \leq F(x) \leq 1,\ x \in (-\infty, \infty)$;

(2) $F(x_1) \leq F(x_2),\ (x_1 < x_2)$;

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

(4) $\lim_{x\to x_0^+}\ F(x) = F(x_0),\ (-\infty < x_0 < \infty)$.

重要公式：$P\{a < X \leq b\} = F(b) - F(a),\ P\{X > a\} = 1 - F(a)$.

3. 定义：如果对于随机变量 X 的分布函数 $F(x)$，存在非负可积函数 $f(x)$，使对于任意实数 x 有 $F\{X \leq x\} = \int_{-\infty}^{x} f(t)\mathrm{d}t$，则称 X 为**连续型随机变量**，称 $f(x)$ 为 X 的**概率密度函数**（Probability Density Function），记为 $X \sim f(x)$.

4. 概率密度的主要性质：

(1) $f(x) \geq 0$;

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

5. 常见连续型随机变量的分布：

(1) 若连续型随机变量 X 有概率密度

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

则称 X 服从区间 (a, b) 上的**均匀分布**（Uniform Distribution），记为 $\boldsymbol{X} \sim U[a, b]$.

(2) 如果连续型随机变量 X 有以下概率密度

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases},$$

其中 $\lambda > 0$，则称 X 服从参数为 λ 的**指数分布**（Exponential Distribution），记为 $\boldsymbol{X} \sim EP(\lambda)$.

其分布函数为

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

(3) 如果连续型随机变量 X 有以下概率密度

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

其中 μ, σ 为常数且 $\sigma > 0$，则称 X 服从参数为 μ, σ 的**正态分布**（Normal Distribution）或**高斯分布**（Gaussian Distribution），简记为 $X \sim N(\mu, \sigma^2)$.

- 称 $N(0,1)$ 为**标准正态分布**，并将其概率密度函数 $\varphi_{0,1}(x)$ 简写为 $\varphi(x)$. 即有

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \ -\infty < x < +\infty.$$

- 标准正态分布的概率密度函数 $\varphi(\cdot)$ 有以下性质:
  (1) 无穷次可微;
  (2) 偶函数;
  (3) 在零点取得最大值;
  (4) 有拐点 ± 1;
  (5) 有水平渐近线（x 轴）.
- 引理: 若 $X \sim N(\mu, \sigma^2)$，则 $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.
- 标准正态分布的分布函数的性质: $\Phi(-x) = 1 - \Phi(x)$.

Lecture 8: Derived Distribution: Single Random Variable

1. 设离散型随机变量 X 的概率分布为

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

令 $Y = g(X)$，则 Y 也是一个离散型随机变量，其分布可按如下步骤求得

(1) 根据函数关系列出 Y 的所有可能值；

(2) 对 Y 的每个可能值 y，$P\{Y = y\}$ 等于所有满足 $g(x_k) = y$ 的 p_k 之和.

2. 对连续型随机变量 X，求 $Y = g(X)$ 的密度函数的基本方法是

(1) 根据函数关系先求 Y 的分布函数

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = \int_{g(x)\leq y} f_X(x)\mathrm{d}x$$

(2) 然后对 $F_Y(y)$ 求导可得 Y 的概率密度.

3. 设随机变量 X 具有概率密度 $f_X(x)$，其中 $-\infty < x < +\infty$，又设函数 $g(x)$ 处处可导，且恒有 $g'(x) > 0$（或 $g'(x) < 0$），则称 $Y = g(x)$ 是连续型随机变量，其概率密度为

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

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

Lecture 9: Multiple Discrete Random Variable

1. 设 (X, Y) 为二维随机向量，对于任意实数 x, y，称二元函数

$$F(x, y) = P\{X \leq x, Y \leq y\}$$

为 (X, Y) 的**联合分布函数**（Joint CDF）。

2. 联合分布函数的主要性质：

(1) $F(x, y)$ 对每个自变量都是广义单增的；

(2) $0 \leq F(x, y) \leq 1$
$F(x, -\infty) = F(-\infty, y) = 0$，$F(-\infty, -\infty) = 0$，$F(+\infty, +\infty) = 1$；

(3) $F(x, y) = F(x + 0, y)$，$F(x, y) = F(x, y + 0)$；

(4) 随机向量 (X, Y) 落在矩形区域 $(x_1, x_2] \times (y_1, y_2]$ 内的概率为

$$\begin{aligned} &P\{x_1 < X \leq x_2, y_1 < Y \leq y_2\} \\ =&F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1). \end{aligned}$$

3. 设二维离散型随机向量 (X, Y) 的所有可能取值为

$$(x_i, y_j), i = 1, 2, \cdots, j = 1, 2, \cdots$$

称

$$\begin{aligned} P\{X = x_i, Y = y_j\} = p_{ij}, \\ i = 1, 2, \cdots, j = 1, 2, \cdots \end{aligned}$$

为该随机向量的**联合分布律**.

二维离散型随机向量的联合分布律也可以用表格表示：

XY	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}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	

5. 设二维离散型随机向量 (X, Y) 的联合分布为

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

则随机变量 X 的**边缘概率分布**（Marginal Probability Distribution）为

$$p_{i\cdot} = P\{X = x_i\} = \sum_j p_{ij}, \qquad i = 1, 2, \cdots$$

而随机变量 Y 的**边缘概率分布**为

$$p_{\cdot j} = P\{Y = y_j\} = \sum_i p_{ij}, \qquad j = 1, 2, \cdots$$

通常将这两个分布分别写在联合分布表右边和下边.

当 $p_{\cdot j} > 0$ 时，称

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

为 $Y = y_j$ 时 X 的**条件概率分布**（Conditional Probability Distribution）.

当 $p_{i\cdot} > 0$ 时，称

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

为 $X = x_i$ 时 Y 的**条件概率分布**.

6. 设随机向量 (X, Y) 的联合分布函数为 $F(x, y)$，边缘分布函数分别为 $F_X(x)$ 和 $F_Y(y)$，若对任意实数 x，y 有

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

则称 X, Y **相互独立**. 若 (X, Y) 是离散型随机向量，X, Y 相互独立的充要条件为 $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}, \quad i, j = 1, 2, \cdots$.

Lecture 10: Multiple Continuous Random Variable

1. 如果存在一个非负函数 $f(x,y)$，使得二维随机向量 (X,Y) 的分布函数 $F(x,y)$ 可以写成

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

则称 (X,Y) 为**二维连续型随机向量**．函数 $f(x,y)$ 称为 (X,Y) 的**联合概率密度**（Joint PDF）．

2. 联合概率密度函数的基本性质：

   1. $f(x,y)\geq 0,\forall x,y$；
   2. $\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\mathrm{d}y=F(\infty,\infty)=1$；
   3. 若函数 f 在点 (x,y) 处连续，则有
   $$\frac{\partial^2F(x,y)}{\partial x\partial y}=f(x,y).$$
   4. 对任意的平面区域 D，
   $$P\{(X,Y)\in D\}=\iint\limits_{(x,y)\in D}f(x,y)\mathrm{d}x\mathrm{d}y.$$

3. 设二维连续型随机向量 (X,Y) 的联合概率密度为 $f(x,y)$，则 X，Y 的**边缘概率密度**分别定义为

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

此时 X，Y 的**边缘分布函数**分别为

$$F_X(x)=\int_{-\infty}^{x}f_X(u)\mathrm{d}u,\quad F_Y(y)=\int_{-\infty}^{y}f_Y(v)\mathrm{d}v.$$

若 $f_Y(y)>0$，在 $Y=y$ 条件下，X 的**条件概率密度**定义为

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

同样地，若 $f_X(x) > 0$ 时，在 $X = x$ 条件下，Y 的**条件概率密度**定义为

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

在 $Y = y$ 条件下，X 的**条件分布函数**定义为

$$\begin{aligned} F_{X|Y}(x|y) &= P\{X \le x|Y = y\} \\ &= \lim_{h\to 0} P\{X \le x|y \le Y \le y+h\} \end{aligned}$$

若 $f(\cdot,\cdot)$ 在点 (x,y) 处连续，$f_Y(\cdot)$ 在点 y 处连续，且 $f_Y(y) > 0$，则

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(s|y)\mathrm{d}s.$$

在 $X = x$ 条件下 Y 的**条件分布函数**定义为

$$\begin{aligned} F_{Y|X}(y|x) &= P\{Y \le y|X = x\} \\ &= \lim_{h\to 0} P\{Y \le y|x \le X \le x+h\} \end{aligned}$$

若 $f(\cdot,\cdot)$ 在点 (x,y) 处连续，$f_X(\cdot)$ 在点 x 处连续，且 $f_X(x) > 0$，则

$$F_{Y|X}(y|x) = \int_{-\infty}^{y} f_{Y|X}(t|x)\mathrm{d}t.$$

4. 设二维随机向量 (X,Y) 的联合分布函数为 $F(x,y)$，边缘分布分别为 $F_X(x)$ 和 $F_Y(y)$，若对任意实数 x,y 有

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

则称 X,Y **相互独立**. 若 (X,Y) 是连续型随机向量，则 X,Y 相互独立的充要条件为：对所有实数 x,y，有 $f(x,y) = f_X(x)\cdot f_Y(y)$。

5. 常用分布：

(1) 设 D 是平面上的有界区域，其面积为 d，若二维随机向量 (X,Y) 的概率密度为

$$f(x,y) = \begin{cases} 1/d, & (x,y) \in D; \\ 0, & \text{otherwise.} \end{cases}$$

则称 (X,Y) 服从 D 上的**均匀分布**.

若 (X,Y) 服从 D 上的均匀分布，则 (X,Y) 落在某一区域 A 内的概率

$$\begin{aligned} P\{(X,Y)\in A\} &= \iint_A f(x,y)\mathrm{d}x\mathrm{d}y \\ &= \iint_{A\cap D} \frac{1}{d}\mathrm{d}x\mathrm{d}y \\ &= \frac{S}{d} \end{aligned}$$

其中 S 为 $A\cap D$ 的面积。

(2) 以下函数为密度的分布称为**二维正态分布**，简记为 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$

$$\begin{aligned} f(x,y) = &\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}, \end{aligned}$$

其中 μ_1,μ_2 为实数，$\sigma_1,\sigma_2>0$，$|\rho|<1$.

设 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，$X$ 和 Y 的边缘概率密度为 $X\sim N(\mu_1,\sigma_1^2), Y\sim N(\mu_2,\sigma_2^2)$. 条件概率密度 $f_{Y|X}(y|x)$ 是一维正态分布，$X=x$ 时 Y 的条件分布为

$$N\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \sigma_2^2(1-\rho^2)\right)$$

X 与 Y 相互独立的充要条件为 $\rho=0$.

Lecture 11: Derived Distribution: Multiple Random Variable

1. 设离散型随机向量 (X,Y) 的联合分布为

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

令 $Z=g(X,Y)$，则 Z 也是一个离散型随机变量，其分布可按如下步骤求得

(1) 根据函数关系列出 Z 的所有可能值；

(2) 对 Z 的每个可能值 z，$P\{Z=z\}$ 等于所有满足 $g(x_i,y_j)=z$ 的 p_{ij} 之和.

2. 对连续型随机向量 (X,Y)，求 $Z=g(X,Y)$ 的密度函数的基本方法是

(1) 根据函数关系先求 Z 的分布函数

$$F_Z(z)=P\{Z\le z\}=P\{g(X,Y)\le z\}$$

(2) 然后对 $F_Z(z)$ 求导可得 Z 的概率密度.

3. $Z=X+Y$
设连续型随机向量 (X,Y) 的联合密度函数为 $f(x,y)$，则 $Z=X+Y$ 的概率密度为

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f(x,z-x)\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} f(z-y,y)\mathrm{d}y. \end{aligned}$$

如果 X 与 Y 相互独立，概率密度分别为 $f_X(x)$ 和 $f_Y(y)$，则 $Z=X+Y$ 的概率密度为

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)\mathrm{d}y. \end{aligned}$$

上述公式称为**卷积公式**.

3. 设 X 与 Y 相互独立，均服从标准正态分布，若 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$，且 X 与 Y 相互独立，则

$$Z = X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

4. $Z = \max\{X, Y\}$，$Z = \min\{X, Y\}$

（相互独立）设随机向量 X 与 Y 相互独立，其分布函数分别为 $F_X(x)$ 和 $F_Y(y)$. 设 $Z = \max\{X, Y\}$ 的分布函数为 $F_{\max}(z)$，则有

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

设 $Z = \min\{X, Y\}$ 的分布函数为 $F_{\min}(z)$，则有

$$\begin{aligned} F_{\min}(z) &= F_X(z) + F_Y(z) - F_X(z)\, F_Y(z) \\ &= 1 - [1 - F_X(z)][1 - F_Y(z)]. \end{aligned}$$

（一般情况）一般地，设连续型随机向量 (X, Y) 的联合分布函数为 $F(x, y)$，则 $U = \max\{X, Y\}$ 的分布为

$$F_U(u) = F(u, u).$$

设其边缘分布函数分别为 $F_X(x)$ 和 $F_Y(y)$，则 $V = \min\{X, Y\}$ 的分布为

$$F_V(v) = F_X(v) + F_Y(v) - F(v, v).$$

$$\begin{aligned} F_V(v) &= P\{\min\{X, Y\} \leq v\} = P\{X \leq v \text{ 或 } Y \leq v\} \\ &= P\{X \leq v\} + P\{Y \leq v\} - P\{X \leq v, Y \leq v\} \\ &= F_X(v) + F_Y(v) - F(v, v) \end{aligned}$$

5. 高维随机向量

- 设 S 是某随机试验的样本空间，$X_1, X_2, \cdots, X_n$ 是该空间上的随机变量，称

$$\vec{X} = (X_1, X_2, \cdots, X_n)$$

为 S 上的或 n **维随机向量**

- 设 $\vec{X}=(X_1,X_2,\cdots,X_n)$ 为 n 维随机向量，称 n 元函数

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

为 $\vec{X}$ 的联合分布函数，其中

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

- 设 $\vec{X}=(X_1,X_2,\cdots,X_n)$ 为 n 维随机向量，如果存在非负函数 $f(\vec{x})$，使得对任意的 $\vec{x}$ 有

$$F(\vec{x})=\int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1}f(\vec{u})\mathrm{d}u_1\mathrm{d}u_2\cdots\mathrm{d}u_n$$

则称 $\vec{X}$ 为 n **维连续型随机向量**，$f(\vec{x})$ 称为 $\vec{X}$ 的**概率密度函数**

- 设 $\vec{X}$ 的联合分布函数为 $F(\vec{x})$,边缘分布分别为 $F_{X_i}(x_i), i=1,2,\cdots,n$,若对任意实数 $x_1,x_2,\cdots,x_n$ 有

$$F(\vec{x})=F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n),$$

则称 $X_1,X_2,\cdots,X_n$ **相互独立**。$X_1,X_2,\cdots,X_n$ 相互独立的充要条件为:对所有的实数 $x_1,x_2,\cdots,x_n$,有 $f(\vec{x})=f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n)$。

6. n 维正态分布：若 $X_1,X_2,\cdots,X_n$ 相互独立，且

$$X_i\sim N(\mu_i,\sigma_i^2),\ i=1,2,\cdots,n$$

则对不全为零的数 $a_1,a_2,\cdots,a_n$,

$$a_1X_1+\cdots+a_nX_n\sim N\left(\sum_{i=1}^n a_i\mu_i,\sum_{i=1}^n a_i^2\sigma_i^2\right).$$

Lecture 12: Expectation and Variance

1. 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 $\sum_k x_k p_k$ 绝对收敛，则称其和为随机变量 X 的**数学期望** (Expectation)，记为 $E(X)$。

2. 设连续型随机变量 X 的概率密度为 $f(x)$，若积分 $\int_{-\infty}^{+\infty} xf(x)dx$ 绝对收敛，则称此积分为随机变量 X 的**数学期望**，记为 $E(X)$。

3. 数学期望的性质: 设 $X, X_1, X_2, \cdots, X_n$ 为随机变量，C 为常数，则有

(1) $E(C)=C$;

(2) $E(CX)=CE(X)$, $E(CX+D)=CE(X)+D$;

(3) $E(X_1+X_2)=E(X_1)+E(X_2)$;
推论：$E\left(\sum_{k=1}^{n} X_k\right)=\sum_{k=1}^{n} E(X_k)$.

(4) 若 X_1, X_2 是相互独立的随机变量，则 $E(X_1X_2)=E(X_1)E(X_2)$.

4. 设 X 为随机变量，$Y=g(X)$，则

(1) 若 X 为离散型随机变量，分布律为

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

则

$$E(Y)=\sum_k g(x_k)p_k.$$

(2) 若 X 为连续型随机变量，概率密度为 $f(x)$，则

$$E(Y)=\int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x.$$

5. 设 (X, Y) 为随机向量，$Z=g(X,Y)$，则

(1) 若 (X,Y) 为离散型随机向量，概率分布为

$$P\{X=x_i,Y=y_j\}=p_{ij},\quad i,j=1,2,\cdots.$$

则

$$E(Z)=\sum_i\sum_j g(x_i,y_j)p_{ij}.$$

(2) 若 (X,Y) 为连续型随机向量，概率密度为 $f(x,y)$，则

$$E(Z)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y.$$

6. 设 X 是一随机变量，若 $[X-E(X)]^2$ 的期望存在，则称该期望为 X 的**方差** (Variance)，记为 $Var(X)$（或 $D(X)$），即

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

称 $\sqrt{Var(X)}$ 为 X 的**标准差** (Standard Deviation)，记为 $\sigma(X)$。
方差的常用计算公式：$Var(X)=E(X^2)-[E(X)]^2$.

7. 设 X、Y 为随机变量，C 为常数，则有

(1) $Var(C)=0$

(2) $Var(X+C)=Var(X)$

(3) $Var(CX)=C^2Var(X)\implies Var(-X)=Var(X)$

(4) $Var(X+Y)=Var(X)+Var(Y)+2E\{[X-E(X)][Y-E(Y)]\}$。
若 X 和 Y 相互独立，则有 $Var(X+Y)=Var(X)+Var(Y)$
性质 4 可以推广到多个相互独立的随机变量。

8. 常用概率分布的期望与方差

随机变量	X	$E(X)$	$Var(X)$
两点分布	$B(1,p)$	p	$p(1-p)$
二项分布	$B(n,p)$	np	$np(1-p)$
泊松分布	$P(\lambda)$	λ	λ
几何分布	$G(p)$	$1/p$	$(1-p)/p^2$
均匀分布	$U[a,b]$	$(a+b)/2$	$(b-a)^2/12$
指数分布	$\mathrm{EP}(\lambda)$	$1/\lambda$	$1/\lambda^2$
正态分布	$N(\mu,\sigma^2)$	μ	σ^2

9. **切比雪夫不等式**: 设随机变量 X 具有数学期望 $E(X)=\mu$, 方差 $D(X)=\sigma^2$，则对于任意正数 ϵ，有不等式

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

Lecture 13: Covariance and Correlation Coefficient

1. 定义: 对于二维随机向量 (X,Y), 称 $Cov(X,Y) := E[(X-EX)(Y-EY)]$ 为 X 与 Y 的**协方差** (Covariance).

2. 设 X,Y,Z 为随机变量，a,b,c,d 为常数，则有

(1) $Cov(X,Y) = Cov(Y,X)$;

(2) $Cov(aX+b, cY+d) = acCov(X,Y)$;

(3) $Cov(X, Y+Z) = Cov(X,Y) + Cov(X,Z)$;

(4) $Cov(X,Y) = E(XY) - E(X) \cdot E(Y)$;

(5) $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X,Y)$. 推论：两随机变量相互独立，则协方差等于零；反之未必成立.

3. 对于二维随机变量 (X,Y)，如果两个变量的方差都不为零，称

$$\rho_{X,Y} := \frac{Cov(X,Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}}$$

为 X 与 Y 的**相关系数** (Correlation Coefficient)，也可以记为 $\rho(X,Y)$.

4. 相关系数表示随机变量之间的**线性相关程度**:

1. $|\rho_{X,Y}| \le 1$.
2. $\rho_{X,Y} = -1$ 当且仅当 $Y = aX + b$，$a < 0$.
3. $\rho_{X,Y} = 1$ 当且仅当 $Y = aX + b$，$a > 0$.

Lecture 14: Moments and Covariance Matrix

1. 定义：对随机变量 X 与正整数 k，

   1. 称 $E(X^k)$ 为 X 的 k 阶原点矩；
   2. 称 $E[(X-EX)^k]$ 为 X 的 k 阶中心矩；
   3. 称 $E(X^kY^l)$ 为 X 和 Y 的 $k+l$ 阶混合矩；
   4. 称 $E\{[(X-EX)^k][(Y-EY)^l]\}$ 为 X 和 Y 的 $k+l$ 阶混合矩。

*2. 定义：对二维随机向量 (X,Y)，称矩阵

$$B=\begin{pmatrix} Cov(X,X) & Cov(X,Y) \\ Cov(Y,X) & Cov(Y,Y) \end{pmatrix}$$

为 (X,Y) 的协方差矩阵.

*3. 定义：设 $\vec{X}=(X_1,X_2,\cdots,X_n)$ 为 n 维随机向量，称矩阵

$$B=(Cov(X_i,X_j))_{n\times n}$$

为 $\vec{X}$ 的协方差矩阵. 协方差阵为对称的半正定矩阵.

*4. 定义：以以下函数为密度的分布称为 n 元正态分布，简记为 $N(\vec{\mu},B)$

$$\varphi(\vec{x})=\frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})^TB^{-1}(\vec{x}-\vec{\mu})\right\},$$

其中 B 为 n 阶正定矩阵，

$$\vec{x}=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},\ \vec{\mu}=\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}.$$

*5. n 元正态分布的性质：若 $\vec{X}\sim N(\vec{\mu},B)$，则

- $\vec{X}$ 的各分量的边缘分布为

$$X_i \sim N(\mu_i, b_{ii}),\ i = 1, 2, \cdots, n$$

  其中 b_{ii} 为 B 的第 i 行第 i 列的元素;

- $\vec{X}$ 的协方差矩阵为 B.

Lecture 15: Law of Large Number and Central Limit Theorem

1. 契比雪夫不等式：设随机变量 X 有期望和方差，则对于任给的 $\varepsilon > 0$，有

$$P(|X-E(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2},$$

即

$$P(|X-E(X)| < \varepsilon) \geq 1-\frac{Var(X)}{\varepsilon^2}.$$

2. **弱大数定律（辛钦大数定律）**：设 $X_1, X_2, \cdots, X_n$ 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 μ，定义 Y_n 为前 n 个随机变量的算术平均，即

$$\boldsymbol{Y}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

则对任意正数 ε，有 $\lim\limits_{n\to\infty} P\left\{|Y_n-\mu| < \varepsilon\right\} = 1$.

3. **伯努利大数定理**：在独立实验序列中，记事件 A 发生的概率为 p。以 $f_n(A)$ 表示前 n 次试验中事件 A 发生的次数，则对任意正数 ε，

$$\mathbf{lim}_{n\to\infty} P\left\{\left|\frac{f_n(A)}{n}-p\right| < \varepsilon\right\} = 1.$$

4. **林德伯格－莱维定理（独立同分布的中心极限定理）**：若 $\{X_n\}_{n\geq 1}$ 独立且同分布，$E(X_i)=\mu$，$Var(X_i)=\sigma^2>0$，令

$$\boldsymbol{Z}_n = \frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

则其分布函数 $F_n(x)$ 收敛到 $\Phi(x)$，即对任何实数 x，

$$\mathbf{lim}_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\{Z_n \leq x\} = \Phi(x)$$

其中 $\Phi(x)$ 是标准正态分布 $N(0,1)$ 的分布函数，

5. **李雅普诺夫定理**：设随机变量序列 $\{X_n\}_{n\geq 1}$ 相互独立，$E(X_i) = \mu_i$，$Var(X_i) = \sigma_i^2 > 0$，记 $B_n^2 = \sum\limits_{i=1}^{n} \sigma_i^2$，若存在正数 δ，使得当 $n \to \infty$ 时，

$$\frac{1}{B_n^{2+\delta}} \sum_{i=1}^{n} E\{|X_i - \mu_i|^{2+\delta}\} \to 0,$$

随机变量之和的标准化变量

$$\boldsymbol{Z}_n = \frac{\sum\limits_{i=1}^{n} X_i - \sum\limits_{i=1}^{n} \mu_i}{B_n}$$

的分布函数 $F_n(x)$ 收敛到 $\Phi(x)$。

6. **棣莫弗－拉普拉斯定理**：设随机变量 X 服从参数为 n, p 的二项分布，即 $X \sim B(n, p)$，则当 n 充分大时，X 近似服从正态分布，即可以近似认为

$$\frac{X - np}{\sqrt{np(1-p)}} \sim N(0, 1).$$

Lecture 16: Sample and Sampling Distribution

1. 称随机变量 $X_1, X_2, \cdots, X_n$ 构成一个**（简单）随机样本** (Random Samplie)，如果这些随机变量相互独立且服从相同的分布。它们共同服从的分布称为**总体分布** (Population Distribution)；样本个数 n 称为**样本容量**.

2. 假设总体 X 服从**离散型**分布 $P\{X = x\} = p(x)$，则 $X_1, X_2, \cdots, X_n$ 的联合分布律为

$$\begin{aligned} & P\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\} \\ =& p(x_1)p(x_2)\cdots p(x_n). \end{aligned}$$

3. 假设总体 X 服从**连续型**分布且密度函数为 $f(x)$，则 $X_1, X_2, \cdots, X_n$ 的联合概率密度为

$$g(x_1, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n).$$

4. 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本，$g(X_1, X_2, \cdots, X_n)$ 是 $X_1, X_2, \cdots, X_n$ 的函数，若 g 中不含未知参数，则称 $g(X_1, X_2, \cdots, X_n)$ 是一个**统计量** (Statistic)。设 $x_1, x_2, \cdots, x_n$ 是相应于 $X_1, X_2, \cdots, X_n$ 的样本值，则称 $g(x_1, x_2, \cdots, x_n)$ 是 $g(X_1, X_2, \cdots, X_n)$ 的**观察值**。

5. 常用统计量:

(1) 对样本 $X_1, X_2, \cdots, X_n$，称

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

为**样本均值**.

(2) 对样本 $X_1, X_2, \cdots, X_n$，称

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

为**样本方差**；称

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

为**样本标准差**.

样本方差的性质:

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

(3) 对样本 $X_1, X_2, \cdots, X_n$ 及正整数 k，称

$$A_k := \frac{1}{n}\sum_{i=1}^{n}X_i^k = \frac{X_1^k + X_2^k + \cdots + X_n^k}{n}$$

为 k **阶样本原点矩**；对 $k \geq 2$，称

$$M_k := \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^k$$

为 k **阶样本中心矩**.

6. 总体分布函数 $F(x)$ 相对应的统计量称为**经验分布函数**。经验分布函数的做法如下: 设 $X_1, X_2, \cdots, X_n$ 是总体 F 的一个样本, 用 $S(x)(-\infty < x < \infty)$ 表示 $X_1, X_2, \cdots, X_n$ 中不大于 x 的随机变量个数, 定义经验分布函数 $F_n(x)$ 为 $F_n(x) = \frac{1}{n}S(x)$, $(-\infty < x < \infty)$。

7. [中心极限定理的常用结论：**大量同分布随机变量的和、平均值近似服从正态分布**。] 设 $X_1, X_2, \cdots, X_n$ 是来自均值为 μ、方差为 σ^2 的总体的简单样本，则当 n 充分大时，近似地有

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

8. 统计量的分布称为**抽样分布** (Sampling Distribution)。以下三个来自正态分布的抽样分布 χ^2 分布，t 分布，F 分布称为**统计学的三大分布**.

9. χ^2 分布

（1）定义：设 $X_1, X_2, \cdots, X_n$ 相互独立，都服从标准正态分布，则

$$Y := \sum_{i=1}^{n} X_i^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

称为服从 n 个自由度的 χ^2 **分布**，记为 $Y \sim \chi_n^2$ 或 $Y \sim \chi^2(n)$.

*n 个自由度的 χ^2 分布的概率密度函数为：

$$f(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0; \\ 0, & x \le 0. \end{cases}$$

（2）χ^2 分布的性质：

1) 若 X 服从标准正态分布，$Y = X^2$，则 Y 服从 1 个自由度的 χ^2 分布，即
$$Y \sim \chi^2(1).$$

2) 可加性：设 $Y_1 \sim \chi^2(m)$, $Y_2 \sim \chi^2(n)$，且两者相互独立，则
$$Y_1 + Y_2 \sim \chi^2(m+n).$$

3) χ^2 分布的数字特征：$E(\chi^2(n)) = n, \quad Var(\chi^2(n)) = 2n.$

（3）对给定的 $\alpha \in (0,1)$，称满足条件 $P\{\chi^2(n) > \chi_\alpha^2(n)\} = \alpha$ 的点 $\chi^2(n)(\alpha)$ 为 $\chi^2(n)$ 分布的**上 α 分位点**.

10. t 分布

（1）定义：设两个随机变量 X, Y 相互独立，并且

$$X \sim N(0,1), \qquad Y \sim \chi^2(n).$$

则称

$$T := \frac{X}{\sqrt{Y/n}}$$

为服从 n 个自由度的 t **分布**，记为 $T \sim t_n$ 或 $T \sim t(n)$。

* 具有 n 个自由度的 t 分布的概率密度函数为:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\cdot\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

注：t 分布的概率密度函数为偶函数.

（2）设 $T \sim t(n)$. 对给定的 $\alpha \in (0,1)$，称满足条件 $P\{T > t_\alpha(n)\} = \alpha$ 的点 $t_\alpha(n)$ 为 t_n 分布的**上 α 分位点**.

设 $Z \sim N(0,1)$，对给定的 $\alpha \in (0,1)$，称满足条件 $P\{Z > Z_\alpha\} = \alpha$ 的点 Z_α 为标准正态分布的**上 α 分位点**.

（3）性质：$t_{1-\alpha}(n) = -t_\alpha(n), \quad Z_{1-\alpha} = -Z_\alpha$.

11. **F 分布**

（1）定义：设两个随机变量 Y_1, Y_2 相互独立，并且

$$Y_1 \sim \chi^2(m), \qquad Y_2 \sim \chi^2(n)$$

则

$$F := \frac{Y_1/m}{Y_2/n} \sim F(m,n).$$

称为自由度为 m 和 n 的 F **分布**，记为 $F \sim F_{m,n}$ 或 $F \sim F(m,n)$.

* 自由度为 m 和 n 的 F 分布的概率密度为:

$$f(x) = \begin{cases} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\cdot\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}, & x > 0; \\ 0, & x \le 0. \end{cases}$$

（2）F 分布的性质:

1) 若 $F \sim F(n,m)$，则 $1/F \sim F(n,m)$.

2) 若 $T \sim t(n)$，则 $T^2 \sim F(1,n)$.

（3）设 $F \sim F(n,m)$．对给定的 $\alpha \in (0,1)$，称满足条件

$$P\{F > F_\alpha(n,m)\} = \alpha$$

的点 $F_\alpha(n,m)$ 为 $F_{m,n}$ 分布的**上 α 分位点**．性质：$F_{1-\alpha}(m,n) = \frac{1}{F_\alpha(n,m)}$.

12. 设 $X_1, X_2, \cdots, X_n$ 是取自正态总体 $N(\mu, \sigma^2)$ 的样本．则 $\overline{X}$ 与 S^2 相互独立，且有

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1), \qquad \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1),$$

$$\frac{\sum_{i=1}^{n}\Big(X_i-\mu\Big)^2}{\sigma^2} \sim \chi^2(n), \qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

13. 设 $X_1, X_2, \cdots, X_m$ 与 $Y_1, Y_2, \cdots, Y_n$ 分别是取自两个相互独立的正态总体

$$N(\mu_1, \sigma_1^2), \qquad N(\mu_2, \sigma_2^2)$$

的样本．则

$$U := \frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} \sim N(0,1),$$

其中 $\overline{X}, \overline{Y}$ 分别是两个样本各自的均值．

14. 设 $X_1, X_2, \cdots, X_m$ 与 $Y_1, Y_2, \cdots, Y_n$ 分别是取自两个相互独立的正态总体

$$N(\mu_1, \sigma^2), \qquad N(\mu_2, \sigma^2)$$

的样本．则

$$T := \frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}} \cdot \sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t_{m+n-2},$$

其中 $\overline{X}$, $\overline{Y}$, S_1^2, S_2^2 分别是两个样本各自的均值及方差.

*15. 设 $X_1, \cdots, X_m$ 与 $Y_1, \cdots, Y_n$ 分别是取自两个相互独立的正态总体

$$N(\mu_1, \sigma_1^2), \qquad N(\mu_2, \sigma_2^2)$$

的样本. 则

$$F := \frac{\sum_{i=1}^{m} \left(X_i - \mu_1\right)^2 \Big/ \left(m\sigma_1^2\right)}{\sum_{j=1}^{n} \left(Y_j - \mu_2\right)^2 \Big/ \left(n\sigma_2^2\right)} \sim F(m, n)$$

16. 设 $X_1, \cdots, X_m$ 与 $Y_1, \cdots, Y_n$ 分别是取自两个相互独立的正态总体

$$N(\mu_1, \sigma_1^2), \qquad N(\mu_2, \sigma_2^2)$$

的样本. 则

$$F := \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(m-1, n-1).$$

其中 S_1^2 和 S_2^2 分别是两个样本各自的方差.

Lecture 17: Point Estimation

1. 设总体 X 的分布函数形式已知，但它的一个或多个参数未知，借助于总体 X 的一个样本来估计总体未知参数的值的问题称为**点估计问题**。

2. 用样本矩来估计总体矩，用样本矩的连续函数来估计总体矩的连续函数，这种估计法称为**矩估计法** (Method of Moments)。具体方法：设总体分布中有 k 个待定参数 $\theta_1, \theta_2, \cdots, \theta_k$。总体 X 的 m 阶原点矩为 $\theta_1, \theta_2, \cdots, \theta_k$ 的函数，记为 $\alpha_m = \alpha_m(\theta_1, \theta_2, \cdots, \theta_k)$。令总体的前 k 阶原点矩分别与同阶的样本原点矩相等，这样就得到了一个 k 元方程组

$$\begin{cases} \alpha_1(\theta_1, \theta_2, \cdots, \theta_k) = A_1 \\ \alpha_2(\theta_1, \theta_2, \cdots, \theta_k) = A_2 \\ \quad \cdots\cdots\cdots \\ \alpha_k(\theta_1, \theta_2, \cdots, \theta_k) = A_k \end{cases}$$

记方程组的解为 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$，用它们作为参数 $\theta_1, \theta_2, \cdots, \theta_k$ 的估计量.

3. 设连续型（离散型）总体 X 的概率密度函数（概率函数）为

$$f(x;\ \theta_1, \theta_2, \cdots, \theta_k).$$

则样本 $X_1, X_2, \cdots, X_n$ 在观测值 $x_1, x_2, \cdots, x_n$ 处的联合概率密度（概率）为

$$L(x_1, \cdots, x_n;\ \theta_1, \cdots, \theta_k) = \prod_{i=1}^{n} f(x_i;\ \theta_1, \cdots, \theta_k).$$

将观测值 $x_1, \cdots, x_n$ 看成固定的，将 L 看做 $\theta_1, \cdots, \theta_k$ 的函数，则该函数被称为**似然函数**（Likelihood Function）。如果似然函数

$$L(x_1, x_2, \cdots, x_n;\ \theta_1, \theta_2, \cdots, \theta_k)$$

在

$$(\theta_1^*, \theta_2^*, \cdots, \theta_k^*)$$

处达到最大值，则称上述参数为未知参数的**极大似然估计**.

(重点) **极大似然估计** (Maximum Likelihood Estimation) 的一般方法:

(1) 写出似然函数 L;

(2) 求似然函数的对数 $\ln L$;

(3) 对 $\ln L$ 求导（偏导）并令导数等于零，得到似然方程组;

(4) 解方程组得到 $\ln L$ 的驻点，判断该驻点是否最大值点.

4. 假设 θ 为总体分布的参数，设 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 是 θ 的一个估计。如果对 θ 的一切可能取值，都有 $E(\hat{\theta}) = \theta$，则称 $\hat{\theta}$ 是 θ 的**无偏估计** (unbiased estimator).
定理：设总体 X 的期望和方差分别为

$$E(X) = \mu, \qquad Var(X) = \sigma^2.$$

从总体取一组样本 $X_1, X_2, \cdots, X_n$，则样本均值 $\overline{X}$ 与样本方差 S^2 分别是 μ 和 σ^2 的无偏估计，即

$$E(\overline{X}) = \mu, \qquad E(S^2) = \sigma^2.$$

5. 设 $\hat{\theta}$ 是参数 θ 的估计量，称

$$MSE(\hat{\theta}) := E[(\hat{\theta} - \theta)^2]$$

为 $\hat{\theta}$ 的**均方误差** (mean squared error)。
均方误差满足以下等式:

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2.$$

当 $\hat{\theta}$ 为 θ 的无偏估计量时，有 $MSE(\hat{\theta}) = Var(\hat{\theta})$。

6. 估计量的比较：在估计量的选取中，

(1) 无偏估计量优于有偏估计量;

(2) 在无偏估计量中，方差越小的越好.

7. 若 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 为参数 θ 的估计量，若对于任意 $\theta \in$，当 $n \to \infty$ 时，$\hat{\theta}(X_1, X_2, \cdots, X_n)$ 依概率收敛于 θ，则称 $\hat{\theta}$ 为 θ 的相合估计量。

Lecture 18: Interval Estimation

1. 设 θ 为总体的待估参数，常数 $\alpha \in (0,1)$，若有两个统计量 $\hat{\theta}_1, \hat{\theta}_2$，使得

$$P\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = 1 - \alpha.$$

则称 $(\hat{\theta}_1, \hat{\theta}_2)$ 为 θ 的**置信系数**为 $1-\alpha$ 的**置信区间**。对给定的置信系数 $1-\alpha$，根据样本观测值确定未知参数 θ 的置信区间 $(\hat{\theta}_1, \hat{\theta}_2)$，称为对参数 θ 的**区间估计**。

2. 总体期望的区间估计（正态总体、方差已知）：设 $X_1, \cdots, X_n$ 是取自正态总体 $N(\mu, \sigma^2)$ 的样本，其中 σ^2 已知．则总体期望 μ 的置信系数为 $1-\alpha$ 的置信区间为

$$\left(\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot Z_{\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot Z_{\alpha/2}\right).$$

3. 总体期望的区间估计（正态总体、方差未知）：设 $X_1, \cdots, X_n$ 是取自正态总体 $N(\mu, \sigma^2)$ 的样本，其中 σ^2 未知。则总体期望 μ 的置信系数为 $1-\alpha$ 的置信区间为

$$\left(\overline{X} - \frac{S}{\sqrt{n}} \cdot t_{\frac{\alpha}{2}}(n-1), \overline{X} + \frac{S}{\sqrt{n}} \cdot t_{\frac{\alpha}{2}}(n-1)\right).$$

总体期望的置信水平为 $1-\alpha$ 的单侧置信区间为 $\left(\overline{X} - \frac{S}{\sqrt{n}} \cdot t_\alpha(n-1), +\infty\right)$.

*4. 总体方差的区间估计（正态总体、期望已知）：设 $X_1, \cdots, X_n$ 是取自正态总体 $N(\mu, \sigma^2)$ 的样本，其中 μ 已知，则总体方差 σ^2 的置信系数为 $1-\alpha$ 的置信区间为

$$\left(\frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{\chi^2_{\alpha/2}(n)}, \frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{\chi^2_{1-\alpha}(n)}\right).$$

*5. 总体方差的区间估计（正态总体、期望未知）：设 $X_1, \cdots, X_n$ 是取自正态总体的样本，则总体方差 σ^2 的置信系数为 $1-\alpha$ 的置信区间为

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right).$$

注：总体标准差 σ 的置信区间为

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}},\ \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}}\right).$$

总体方差的置信水平为 $1-\alpha$ 的单侧置信区间为：$\left(0, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$

*6. 两个正态总体的期望差的区间估计（两个正态总体，方差已知）：设两个总体的方差 σ_1^2 与 σ_2^2 均已知，则总体期望的差 $\mu_1-\mu_2$ 的置信系数为 $1-\alpha$ 的置信区间为

$$\left(\overline{X}-\overline{Y}-\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}\cdot Z_{\alpha/2},\ \overline{X}-\overline{Y}+\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}\cdot Z_{\alpha/2}\right)$$

*7. 两个正态总体的期望差的区间估计（两个正态总体，方差未知，但相等）：设两个总体的方差未知，但知道它们相等，则总体期望的差 $\mu_1-\mu_2$ 的置信系数为 $1-\alpha$ 的置信区间为

$$\left(\overline{X}-\overline{Y}-S\cdot\sqrt{\frac{1}{m}+\frac{1}{n}}\cdot t_{\alpha/2}(m+n-2), \overline{X}-\overline{Y}+S\cdot\sqrt{\frac{1}{m}+\frac{1}{n}}\cdot t_{\alpha/2}(m+n-2)\right),$$

其中 $S=\sqrt{\dfrac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}}$.

*8. 方差比的区间估计（两个正态总体，期望已知）：设两个总体的期望 μ_1 与 μ_2 已知，则两个总体的方差比 σ_1^2/σ_2^2 的置信系数为 $1-\alpha$ 的置信区间为

$$\left(\frac{\sum\limits_{i=1}^{m}\left(X_i-\mu_1\right)^2\Big/m}{F_{\alpha/2}(m,n)\sum\limits_{j=1}^{n}\left(Y_j-\mu_1\right)^2\Big/n},\ \frac{\sum\limits_{i=1}^{m}\left(X_i-\mu_2\right)^2\Big/m}{F_{1-\alpha/2}(m,n)\sum\limits_{j=1}^{n}\left(Y_j-\mu_2\right)^2\Big/n}\right).$$

*9. 方差比的区间估计（两个正态总体，期望未知）：设两个总体的期望 μ_1 与 μ_2 未知，则两个总体的方差比 σ_1^2/σ_2^2 的置信系数为 $1-\alpha$ 的置信区间为

$$\left(\frac{S_1^2}{S_2^2\,F_{\alpha/2}(m-1,n-1)},\ \frac{S_1^2}{S_2^2\,F_{1-\alpha/2}(m-1,n-1)}\right).$$

Lecture 19: Hypothesis Testing

1. 参数的假设检验主要用置信区间的方法，步骤如下：

(1) 根据实际问题，提出原假设 H_0 与备择假设 H_1；

(2) 假定 H_0 成立，选择分布已知的统计量；

(3) 由给定的显著性水平 α，计算统计量的临界值，由此划分拒绝域与接受域；

(4) 根据样本观测值，计算统计量的观测值；

(5) 检查观测值所在区域，并作出结论。

- 如果落在拒绝域，则拒绝原假设 H_0
- 如果落在接受域，则接受原假设 H_0

2. **第一类错误（弃真错误）**：原假设 H_0 符合实际情况，而检验结果把它否定了。$P\{reject\ H_0|H_0\ is\ true\} = \alpha$。**第二类错误（取伪错误）**：原假设 H_0 不符合实际情况，而检验结果把它肯定了。$P\{accept\ H_0|H_0\ is\ false\} = \beta$。

3. 单个正态总体，方差 σ^2 已知

条件	单个正态总体，方差 σ^2 已知		
原假设	$H_0 : \mu = \mu_0$		
备择假设	$\mu < \mu_0$	$\mu \neq \mu_0$	$\mu > \mu_0$
统计量	$Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$		
拒绝域	$Z \leq -Z_\alpha$	$\|Z\| \geq Z_{\frac{\alpha}{2}}$	$Z \geq Z_\alpha$

4. 单个正态总体，方差 σ^2 未知

条件	单个正态总体，方差未知		
原假设	$H_0: \mu = \mu_0$		
备择假设	$\mu < \mu_0$	$\mu \neq \mu_0$	$\mu > \mu_0$
统计量	$T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$		
拒绝域	$T \leq -t_\alpha(n-1)$	$\|T\| \geq t_{\frac{\alpha}{2}}(n-1)$	$T \geq t_\alpha(n-1)$

5. 单个正态总体的方差

检验目标	单个正态总体的方差		
原假设	$H_0: \sigma^2 = \sigma_0^2$		
备择假设	$\sigma^2 < \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\sigma^2 > \sigma_0^2$
统计量	$\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$		
拒绝域	$\chi^2 \leq \chi^2_{1-\alpha}(n-1)$	$\chi^2 \geq \chi^2_{\frac{\alpha}{2}}(n-1)$ 或 $\chi^2 \leq \chi^2_{1-\frac{\alpha}{2}}(n-1)$	$\chi^2 \geq \chi^2_\alpha(n-1)$

*6. 两个正态总体，方差 σ_1^2, σ_2^2 已知

条件	两个正态总体，方差 σ_1^2, σ_2^2 已知		
原假设	$H_0: \mu_1 = \mu_2$		
备择假设	$\mu_1 < \mu_2$	$\mu_1 \neq \mu_2$	$\mu_1 > \mu_2$
统计量	$Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim N(0,1)$		
拒绝域	$Z \leq -Z_\alpha$	$\|Z\| \geq Z_{\frac{\alpha}{2}}$	$Z \geq Z_\alpha$

*7. 两个正态总体，方差未知但相等

条件	两个正态总体，方差未知但相等		
原假设	$H_0: \mu_1 = \mu_2$		
备择假设	$\mu_1 < \mu_2$	$\mu_1 \neq \mu_2$	$\mu_1 > \mu_2$
统计量	$T = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}} \cdot \sqrt{\frac{1}{m} + \frac{1}{n}}}$		
拒绝域	$T \leq -t_\alpha(m+n-2)$	$\|T\| \geq t_{\frac{\alpha}{2}}(m+n-2)$	$T \geq t_\alpha(m+n-2)$

*8. 两个正态总体的方差

检验目标	两个正态总体的方差		
原假设	$H_0: \sigma_1^2 = \sigma_2^2$		
备择假设	$\sigma_1^2 < \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$
统计量	$F = \dfrac{S_1^2}{S_2^2} \sim F(m-1, n-1)$		
拒绝域	$F \leq F_{1-\alpha}(m-1, n-1)$	$F \geq F_{\frac{\alpha}{2}}(m-1, n-1)$ 或 $F \leq F_{1-\frac{\alpha}{2}}(m-1, n-1)$	$F \geq F_\alpha(m-1, n-1)$