概率论与数理逻辑笔记整理V2.0

陈鸿峥

2018.11 *

目录

1 基本概念 **2**

1.1 事件与概率 2

1.2 条件概率 2

2 随机变量及其分布 **3**

2.1 基本概念 3

2.2 随机变量的函数的分布 5

2.3 常见的离散分布 5

2.4 常见的连续分布 7

3 多维随机变量及其分布 **9**

3.1 边缘分布 9

3.2 随机变量的函数的分布 10

3.3 数字特征 11

4 大数定律 **12**

4.1 大数定律 12

4.2 中心极限定理 13

5 统计 **14**

5.1 样本与抽样 14

5.2 抽样分布 14

5.3 参数估计 15

5.4 可视化 16

*Build 20181105

1　基本概念

1.1　事件与概率

命题 1. 事件的基本运算

1. 分配律：$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$, $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$
2. 德摩根律：$\overline{A\cup B}=\overline{A}\cap\overline{B}$, $\overline{A\cap B}=\overline{A}\cup\overline{B}$

注意概率是对一个集合的函数，有如下定义.

定义 1 (概率). 随机试验E的所有可能结果构成E的样本空间Ω，Ω的子集称为事件，Ω的幂集构成E的事件空间$\mathcal{F}$，记概率函数$\mathbb{P}(\cdot):\mathcal{F}\mapsto\mathbb{R}$满足：

1. 非负性：$\mathbb{P}(A)\geq 0,\forall A\in\mathcal{F}$
2. 规范性：$\mathbb{P}(\Omega)=1$
3. 可列可加性：$A_1,A_2,\ldots$为两两不相容的事件，$\mathbb{P}\left(\bigcup\limits_{i=1}^{\infty}A_i\right)=\sum\limits_{i=1}^{\infty}\mathbb{P}(A_i)$

由定义可得概率一些基本性质：

1. $\mathbb{P}(\varnothing)=0$, $\mathbb{P}(A)\leq 1$
2. 有限可加性：$A_1,A_2,\ldots$两两不相容，$\mathbb{P}\left(\bigcup\limits_{i=1}^{n}A_i\right)=\sum\limits_{i=1}^{n}\mathbb{P}(A_i)$
3. 若$A\subset B$，则$\mathbb{P}(B-A)=\mathbb{P}(B)-\mathbb{P}(A)$, $\mathbb{P}(B)\geq\mathbb{P}(A)$
4. 逆事件概率：$\mathbb{P}\left(\overline{A}\right)=1-\mathbb{P}(A)$
5. 容斥原理：

$$\mathbb{P}\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{k=1}^{n}\left((-1)^{k-1}\sum_{\substack{I\subset\{1,\ldots,n\}\\|I|=k}}\mathbb{P}\left(\bigcap_{i\in I}A_i\right)\right)$$

本章的重点在于组合计数，正确计算事件数目并套用相应的公式即可.

1.2　条件概率

定义 2 (条件概率). 设A,B为两个事件，且$\mathbb{P}(A)>0$，则称

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

为在事件A发生的条件下事件B发生的条件概率

进而有

$$\mathbb{P}(B\mid A)\mathbb{P}(A)=\mathbb{P}(AB)=\mathbb{P}(A\mid B)\mathbb{P}(B)$$

定理 1 (乘法公式). 设$A_1, A_2, \ldots, A_n$为n个事件，$n \geq 2$，且$\mathbb{P}(A_1 A_2 \cdots A_{n-1}) > 0$，则

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

定义 3 (划分). 两两交为空，所有并为全集

定理 2 (全概率公式). 设试验E的样本空间为S，A为E的事件，$B_1, B_2, \ldots, B_n$为S的一个划分，且$\mathbb{P}(B_i) > 0$，则

$$\mathbb{P}(A) = \mathbb{P}(AB_1) + \mathbb{P}(AB_2) + \cdots + \mathbb{P}(AB_n) = \mathbb{P}(A \mid B_1) \mathbb{P}(B_1) + \mathbb{P}(A \mid B_2) + \cdots + \mathbb{P}(A \mid B_n) \mathbb{P}(B_n)$$

定理 3 (贝叶斯(Bayes)公式). 设试验E的样本空间为S，A为E的事件，$B_1, B_2, \ldots, B_n$为S的一个划分，且$\mathbb{P}(A) > 0, \mathbb{P}(B_i) > 0$，则

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(B_i A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_i) \mathbb{P}(B_i)}{\sum_{i=1}^{n} P(A \mid B_i) \mathbb{P}(B_i)}$$

特别地，当$n = 2$时有

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B) \mathbb{P}(B)}{\mathbb{P}(A \mid B) \mathbb{P}(B) + \mathbb{P}\left(A \mid \overline{B}\right) \mathbb{P}\left(\overline{B}\right)}$$

注意贝叶斯公式是用**先验概率**推**后验概率**.

在计算条件概率时一定要注意前提条件是什么，并将题设进行转换.

定义 4 (独立性). 对于事件$A_1, \ldots, A_n$，

- 若$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \mathbb{P}(A_j), \forall i, j$，则称$\{A_1, \ldots, A_n\}$两两(pairwise)独立
- 若$\mathbb{P}\left(\bigcap_{j \in I} A_j\right) = \prod_{j \in I} \mathbb{P}(A_j), \forall I \in 2^{[n]}$，其中$2^{[n]}$为$\{A_i\}_{i=1}^{n}$的所有子集，则称$\{A_1, \ldots, A_n\}$相互(mutually)独立

区分以下两个概念

1. A, B对立(exclusive)/不相容 $\Leftrightarrow \mathbb{P}(A \cap B) = 0$，即不相交(disjoint)
2. A, B独立(independent) $\Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$，即不相关(unrelated)

2 随机变量及其分布

2.1 基本概念

定义 5. 对于离散随机变量X，其概率质量函数*(PMF)*为$f_X(k) = \mathbb{P}(X = k)$，分布函数为$F_X(x) = \mathbb{P}(X \leq x) = \sum_{k \leq x} f_X(k)$

定义 6. *对于连续随机变量X，其累积密度函数(CDF)为$F_X(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f_X(z)\,\mathrm{d}z\, \int_{-\infty}^{x} f(t)\,\mathrm{d}x$，其中$f_X(x)$为$X$的概率密度函数(PDF)，也即$f_X(x) = \dfrac{\mathrm{d}F_X(x)}{\mathrm{d}x}$. 一定要注意，$f_X(x) \neq \mathbb{P}(X = x)$!*

注意积分区间！注意要写变量范围！

定义 7 (期望). *设Y是随机变量X的连续函数$Y = g(X)$*

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_x x f_X(x) & \mathbb{E}(g(X)) &= \sum_x g(x) f_X(x) \\
\mathbb{E}(X) &= \int_{-\infty}^{+\infty} x f_X(x)\,\mathrm{d}x & \mathbb{E}(g(X)) &= \int_{-\infty}^{+\infty} g(x) f_X(x)\,\mathrm{d}x
\end{aligned}
$$

期望具有线性性，即

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y), \mathbb{E}(cX) = c\mathbb{E}(X)$$

若X, Y相互独立，则

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

但反过来不能推相互独立

定义 8 (方差).

$$
\begin{aligned}
\mathbb{D}(X) = \mathrm{Var}(X) = \sigma^2 = \mathbb{E}\left((X-\mathbb{E}(X))^2\right) &= \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 \ge 0 \\
&= \sum_{k=1}^{\infty} [x_k - \mathbb{E}(X)]^2 p_k \\
&= \int_{-\infty}^{\infty} [X - \mathbb{E}(X)]^2 f(x)\,\mathrm{d}x
\end{aligned}
$$

标准差或均方差则是σ

一般通过求$\mathbb{E}(X)$和$\mathbb{E}\left(X^2\right)$来求方差
由方差定义和期望的线性性有

$$\mathbb{D}(aX+b) = a^2\mathbb{D}(X)$$

注意方差并不是线性的

$$
\begin{aligned}
\mathbb{D}(X+Y) &= \mathbb{E}\left((X+Y)^2\right) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\
&= \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 + \mathbb{E}\left(Y^2\right) - \mathbb{E}(Y)^2 + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)) \\
&= \mathbb{D}(X) + \mathbb{D}(Y) + 2\mathrm{Cov}(X,Y)
\end{aligned}
$$

定义 9 (上α分位点).

$$\mathbb{P}(X > z_\alpha) = \alpha,\ \alpha \in (0,1)$$

2.2 随机变量的函数的分布

定理 4. 若X为连续型随机变量，g为单调递增函数（反函数存在且单调递增），且$Y=g(X)$，那么

$$f_Y(y)=f_X(g^{-1}(y))(g^{-1})'(y)$$

特别地，对于$Y=X^2$

$$f_Y(y)=\frac{1}{2\sqrt{y}}[f_X(\sqrt{y})+f_X(-\sqrt{y})],\, y>0$$

一般情况则先求Y的概率分布函数$F_Y(y)$，然后对$F_Y(y)$求导

2.3 常见的离散分布

1. 伯努利/两点/0-1分布 **Bernoulli(p)**（二项分布的特殊情形）

$$f_X(k)=\begin{cases}1-p & k=0\\ p & k=1\end{cases}$$
$$\mathbb{E}(X)=p$$
$$\mathbb{D}\,(X)=p(1-p)$$

2. 二项分布 **Binomial(n,p)**

$$f_X(k)=\binom{n}{k}p^k(1-p)^{n-k},\, 0\le k\le n$$
$$\mathbb{E}(X)=n\cdot p$$
$$\mathbb{D}\,(X)=np(1-p)$$

   e.g. 扔n次硬币扔到k次正面（做实验n次，记录成功的次数）

3. 几何分布 **Geometric(p)**（负二项分布的特殊情形）

$$f_X(k)=(1-p)^k\cdot p, k\ge 0$$
$$\mathbb{E}(X)=\frac{1-p}{p}$$

   e.g. 扔k次反面直至扔到正面（做实验直到你成功，记录失败的次数）

4. 负二项分布 **NegetiveBinomial(r,p)**

$$f_X(k)=\binom{k+r-1}{r-1}p^r(1-p)^k,\, k\ge 0$$
$$\mathbb{E}(X)=r\cdot\frac{1-p}{p}$$

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{\infty}\binom{k+r-1}{r-1}p^r(1-p)^k \\
&= \sum_{k=0}^{\infty} r\left(\binom{k+r}{r}-\binom{k+r-1}{r-1}\right)p^r(1-p)^k \qquad k\binom{n}{k}=n\binom{n-1}{k-1}\text{的变形} \\
&= rp^r\left(\sum_{k=0}^{\infty}\binom{k+r}{r}(1-p)^k-\sum_{k=0}^{\infty}\binom{k+r-1}{r-1}(1-p)^k\right) \\
&= rp^r\left(\sum_{k=0}^{\infty}\binom{k+r}{k}(1-p)^k-\sum_{k=0}^{\infty}\binom{k+r-1}{k}(1-p)^k\right) \\
&= rp^r\left(\sum_{k=0}^{\infty}(-1)^k\binom{-r-1}{k}(1-p)^k-\sum_{k=0}^{\infty}(-1)^k\binom{-r}{k}(1-p)^k\right) \\
&\quad \text{这步是关键，将变化的}(k+r)\text{转成}(-r-1)\text{，使得可以正常使用二项式定理} \\
&= rp^r\left(\sum_{k=0}^{\infty}\binom{-r-1}{k}1^{-r-1-k}(p-1)^k-\sum_{k=0}^{\infty}\binom{-r}{k}1^{-r-k}(p-1)^k\right) \\
&= rp^r\left((1+(p-1))^{-r-1}-(1+(p-1))^{-r}\right) \qquad \text{牛顿二项式} \\
&= r(p^{-1}-1) \\
&= r\frac{1-p}{p}
\end{aligned}
$$

补充证明：

$$
\begin{aligned}
\binom{k+r}{k} &= \frac{(k+r)(k+r-1)\cdots(r+1)}{k(k-1)\cdots 1} \\
&= (-1)^k\frac{(-k-r)(-k-r-1)\cdots(-r-1)}{k(k-1)\cdots 1} \\
&= (-1)^k\frac{(-r-1)(-r-2)\cdots(-r-1-k+1)}{k(k-1)\cdots 1} \qquad \text{把分子各项逆过来} \\
&= (-1)^k\binom{-r-1}{k}
\end{aligned}
$$

e.g. 扔k次反面直到有r个正面（做实验直到你获得r次成功，记录失败次数）

5. 超几何分布 **HyperGeometric(N,n,M)**

$$
\begin{aligned}
f_X(k) &= \frac{\binom{M}{k}\cdot\binom{N-M}{n-k}}{\binom{N}{n}} \\
\mathbb{E}(X) &= n\frac{M}{N}
\end{aligned}
$$

e.g. M个产品中有N个次品，检查n次得到k个次品

6. 泊松分布 **Poisson(λ),**$\lambda > 0$

$$f_X(k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda},\ \lambda > 0, k \geq 0$$
$$\mathbb{E}(X) = \lambda$$

$X \sim B(n,p)$，若$p = \dfrac{\lambda}{n}$，且n非常大，则

$$\begin{aligned}\mathbb{P}\left(X = k\right) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!}\frac{\lambda^k}{n^k}\left(1 + \frac{-\lambda}{n}\right)^{n-k} \\ &\approx \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}\end{aligned}$$

一般$n \geq 20, p \leq 0.05$时，即可用近似

$$\begin{aligned}\mathbb{E}\left(X\right) &= \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} \\ &= \lambda\mathrm{e}^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = \lambda\end{aligned}$$

注意泊松分布具有无记忆性(memoryless)，即

$$\mathbb{P}\left(X \geq a \mid X \geq b\right) = \mathbb{P}\left(X \geq a - b\right)$$

2.4 常见的连续分布

1. 均匀分布 **Uniform(a,b),**$a < b$

$$f_X(x) = \frac{1}{b-a},\ a \leq x \leq b$$
$$\mathbb{E}\left(X\right) = \frac{a+b}{2}$$
$$\mathbb{D}\left(X\right) = \frac{(b-a)^2}{12}$$

2. 指数分布 **Exponential(θ),**$\theta > 0$

$$f_X(x) = \frac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}},\ x > 0$$
$$\mathbb{E}\left(X\right) = \theta = \frac{1}{\lambda}$$

与泊松分布类似，同样具有无记忆性

3. 正态分布 **Normal**(μ, σ^2)

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{\frac{-(x-\mu)^2}{2\sigma^2}}$$
$$\mathbb{E}(X) = \mu$$

标准正态分布$N(0,1)$

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

算平方

$$\begin{aligned}
I^2 &= \left(\int_{-\infty}^{+\infty} \mathrm{e}^{\frac{-x^2}{2}}\ \mathrm{d}x\right)^2 \\
&= \int_{-\infty}^{+\infty} \mathrm{e}^{\frac{-x^2}{2}}\ \mathrm{d}x \int_{-\infty}^{+\infty} \mathrm{e}^{\frac{-y^2}{2}}\ \mathrm{d}y \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{e}^{\frac{-(x^2+y^2)}{2}}\ \mathrm{d}x\,\mathrm{d}y \\
&= \int_{0}^{2\pi}\int_{0}^{+\infty} r\mathrm{e}^{\frac{-r^2}{2}}\ \mathrm{d}r\ \mathrm{d}\theta \\
&\quad \text{重积分极坐标变量代换 } \mathrm{d}S = r\,\mathrm{d}r\,\mathrm{d}\theta \\
&= -\int_{0}^{2\pi}\int_{0}^{+\infty} \mathrm{d}\mathrm{e}^{\frac{-r^2}{2}}\ \mathrm{d}\theta \\
&= -\int_{0}^{2\pi} (0-1)\ \mathrm{d}\theta \\
&= \int_{0}^{2\pi} 1\ \mathrm{d}\theta = 2\pi
\end{aligned}$$

也即概率积分$I = \sqrt{2\pi}$

若$X \sim N(\mu, \sigma^2)$，则$Y = aX + b \sim N(a\mu + b, (a\sigma)^2)$，特别地，

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

若$X_i \sim N(\mu_i, \sigma_i^2)$相互独立，则它们的和

$$Z = X_1 + X_2 + \cdots + X_n \sim N(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$$

若$X \sim N(0,1)$，且$Y = X^2$，则

$$f_Y(y) = \frac{1}{\sqrt{2\pi}}y^{-1/2}\mathrm{e}^{-y/2} \implies Y \sim \Gamma(\frac{1}{2}, 2)$$

4. 伽马分布 $\mathbf{Gamma}(\alpha,\beta)\equiv\Gamma(\alpha,\beta)\equiv\Gamma(k,\theta)\,,k=\alpha,\theta=1/\beta$

$$
\begin{aligned}
f_X(x;\alpha,\beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\beta x},\ x,\alpha,\beta>0 \\
f_X(x;k,\theta) &= \frac{1}{\Gamma(k)\theta^k}x^{k-1}\mathrm{e}^{-\frac{x}{\theta}},\ x,k,\theta>0 \\
F(x;k,\theta) &= \int_0^x f(u;k,\theta)\,\mathrm{d}u = \frac{\gamma\left(k,\frac{x}{\theta}\right)}{\Gamma(k)} \\
\mathbb{E}\left(X\right) &= \frac{\alpha}{\beta} = k\theta \\
\mathbb{D}\left(X\right) &= \frac{\alpha}{\beta^2} = k\theta^2
\end{aligned}
$$

若$X_i\sim\Gamma(k_i,\theta)$相互独立，则它们的和

$$
Z=\sum_{i=1}^N X_i\sim\Gamma\left(\sum_{i=1}^n k_i,\theta\right)
$$

注：通过$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$证明

3 多维随机变量及其分布

3.1 边缘分布

定义 10. *二维随机变量(X,Y)的分布函数/联合(joint)分布函数定义如下*

$$
F(x,y)=\mathbb{P}\left(X\le x,Y\le y\right)=\sum_{x_i\le x}\sum_{y_i\le y}p_{ij}=\int_{-\infty}^{y}\int_{-\infty}^{x}f(u,v)\ \mathrm{d}u\ \mathrm{d}v
$$

其中，$f(x,y)$为X,Y的联合密度函数

进而，对于离散型随机变量变量有，

$$
\mathbb{P}\left(x_1<X\le x_2,y_1<Y\le y_2\right)=F(x_2,y_2)-F(x_2,y_1)+F(x_1,y_1)-F(x_1,y_2)
$$

连续型随机变量有，

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

若$f(x,y)$在(x,y)连续，则$\dfrac{\partial^2F(x,y)}{\partial x\partial y}=f(x,y)$

定义 11 (边缘(marginal)分布)**.**

$$
\begin{aligned}
F_X(x) &= \mathbb{P}\left(X \leq x, Y < \infty\right) = \lim_{y \to \infty} F(x, y) = F(x, \infty) \\
&= \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} f(x, y) \ \mathrm{d}y \right] \mathrm{d}x \\
&= \int_{-\infty}^{x} f_X(x) \ \mathrm{d}x \\
f_X(x) &= \int_{-\infty}^{\infty} f(x, y) \ \mathrm{d}y
\end{aligned}
$$

注意积分区间不一定是从负无穷到正无穷，而是概率有$(0,1)$之间值的区域
对于二元随机变量的概率密度和概率分布函数有如下关系

$$
\begin{tikzcd}
f(x,y) \arrow[r, "\text{Int } y"] \arrow[d, "\text{Int } x"] \arrow[ddrr, "\text{Int } x\ y"] & f_X(x) \arrow[r, "\text{Int } x"] & F_X(x) \\
f_Y(y) \arrow[d, "\text{Int } y"] & & \\
F_Y(y) & & F(x,y) \arrow[ll, "x \to \infty"'] \arrow[uu, "y \to \infty"']
\end{tikzcd}
$$

定义 12 (条件概率密度与分布函数)**.**

$$
\begin{aligned}
f_{X|Y}(x \mid y) &= \frac{f(x,y)}{f_Y(y)} \\
F_{X|Y}(x \mid y) &= \mathbb{P}\left(X \leq x \mid Y = y\right) = \int_{-\infty}^{x} \frac{f(x,y)}{f_Y(y)} \ \mathrm{d}x
\end{aligned}
$$

定义 13 (相互独立)**.**

$$
\begin{aligned}
F(x,y) &= F_X(x) F_Y(y) \\
f(x,y) &= f_X(x) f_Y(y)
\end{aligned}
$$

3.2 随机变量的函数的分布

均通过求$F_Z(z) = \mathbb{P}\left(z \leq Z\right)$交换积分次序得到

3.2.1 $Z = X + Y$

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

若X和Y相互独立，则有卷积公式

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

3.2.2 $Z = X/Y$

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

3.2.3 $Z = XY$

$$f_{XY}(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f(x, \frac{z}{x})\,\mathrm{d}x$$

3.2.4 $Z = \max\{X_1, \ldots, X_n\}$

$$F_{\max}(z) = F_{X_1}(z) F_{X_2}(z) \cdots F_{X_n}(z)$$

3.2.5 $Z = \min\{X_1, \ldots, X_n\}$

$$F_{\min}(z) = 1 - [1 - F_{X_1}(z)][1 - F_{X_2}(z)] \cdots [1 - F_{X_n}(z)]$$

3.2.6 函数分布

若Z是随机变量X, Y的连续函数$Z = g(X, Y)$，则

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

根据这条式子，方差、协方差等等均可以直接通过积分计算

3.3 数字特征

定义 14 (协方差)**.**

$$\mathrm{Cov}\left(X, Y\right) = \mathbb{E}\left([X - \mathbb{E}\left(X\right)][Y - \mathbb{E}\left(Y\right)]\right) = \mathbb{E}\left(XY\right) - \mathbb{E}\left(X\right)\mathbb{E}\left(Y\right)$$

协方差的性质：

$$\mathrm{Cov}\left(aX, bY\right) = ab\mathrm{Cov}\left(X, Y\right)$$
$$\mathrm{Cov}\left(X_1 + X_2, Y\right) = \mathrm{Cov}\left(X_1, Y\right) + \mathrm{Cov}\left(X_2, Y\right)$$

定义 15 (相关系数)**.** 用来表征X, Y线性关系的量

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

$|\rho_{XY}|$较大，均方误差小，线性关系强； $\rho_{XY} \leq 1$，取等的充要条件为$\exists a, b$使得$\mathbb{P}\left(Y = a + bX\right) = 1$

相关性与独立性没有必然联系：相关性是对线性关系来说的，而独立性是对一般关系来说的

定义 16 (矩)**.** 设X, Y为随机变量，若$\mathbb{E}\left(X^k\right)$存在，则称它为$X$的$k$阶*(原点)*矩，称$\mathbb{E}\left([X-\mathbb{E}(X)]^k\right)$为$k$阶中心矩，称$\mathbb{E}\left(X^kY^l\right)$为$k+l$阶混合矩，称$\mathbb{E}\left([X-\mathbb{E}(X)]^k[Y-\mathbb{E}(Y)]^l\right)$为$k+l$阶混合中心矩

定义 17 (协方差矩阵)**.** 设n维随机变量$(X_1, X_2, \ldots, X_n)$的二阶混合中心距

$$c_{ij} = \mathrm{Cov}\left(X_i, X_j\right) = \mathbb{E}\left([X_i - \mathbb{E}(X_i)][X_j - \mathbb{E}(X_j)]\right)$$

都存在，则协方差矩阵

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

又$c_{ij} = c_{ji}$，故上述矩阵是个对称矩阵

定理 5. 若$X_1, \ldots, X_n$独立，则

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}(X_i)$$
$$\mathbb{D}\left(\prod_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{D}(X_i)$$
$$\mathrm{Cov}\left(X_i, X_j\right) = 0,\ i \neq j$$

$Y = g(X)$的随机变量分布一般通过求分布函数后求导得到其概率密度函数

4 大数定律

4.1 大数定律

定理 6 (切比雪夫(Chebyshev)不等式)**.** 设随机变量X的数学期望$\mathbb{E}(X) = \mu$，方差$\mathbb{D}(X) = \sigma^2$

$$\mathbb{P}\left(|X-\mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2}$$

或

$$\mathbb{P}\left(|X-\mu| < \varepsilon\right) \geq \frac{\sigma^2}{\varepsilon^2}$$

定理 7 (弱大数定律(辛钦))**.** 若$X_1, X_2, \ldots$为独立随机变量且同等分布*(iid)*，$\mathbb{E}(X_i) = \mu, \mathbb{D}(X_i) = \sigma^2$，则

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right) = 0, \forall\varepsilon > 0$$

或

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \varepsilon\right) = 1, \forall\varepsilon > 0$$

可记成$\bar{X} \xrightarrow{P} \mu$

定理 8 (实际推断原理). 概率很小的事件在一次试验中实际上几乎是不发生的

定理 9 (强大数定律). 若$X_1, X_2, \ldots$为独立随机变量且同等分布，$\mathbb{E}(X_i) = \mu, \mathbb{D}(X_i) = \sigma^2$，则

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\right) = 1$$

4.2 中心极限定理

定义 18 (标准化变量). 若随机变量X的均值为μ，方差为σ^2，则X的标准化变量为

$$Z = \frac{X-\mu}{\sigma}$$

有$\mathbb{E}(Z) = 0$，$\mathbb{D}(Z) = 1$

定理 10 (独立同分布的中心极限定理). 若$X_1, X_2, \ldots$为独立随机变量且同等分布，且$\mathbb{E}(X_k) = \mu$，$\mathbb{D}(X_k) = \sigma^2 > 0$，则随机变量之和的标准化变量

$$Y_n = \frac{\sum_{k=1}^{n} X_k - \mathbb{E}\left(\sum_{k=1}^{n} X_k\right)}{\sqrt{\mathbb{D}\left(\sum_{k=1}^{n} X_k\right)}} = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n\sigma^2}}$$

的分布函数$F_n(x)$对任意x满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} \mathbb{P}(Y_n \leq x) = \int_{-\infty}^{x} \frac{1}{2\pi}\mathrm{e}^{-t^2/2}\,\mathrm{d}t = \Phi(x)$$

也即，近似地

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

努力将原变量转化为标准化变量形式，以使用标准正态分布解题

推论 (棣莫弗-拉普拉斯(De Moivre-Laplace)定理). 设随机变量$\eta_1, \eta_2, \ldots$服从参数为$n, p(0 < p < 1)$的二项分布，则对于任意x，有

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\eta_n - np}{\sqrt{np(1-p)}} \leq x\right) = \Phi(x)$$

定理 11 (李雅普诺夫(Lyapunov)定理). 若$X_1, X_2, \ldots$为独立随机变量且同等分布，且$\mathbb{E}(X_k) = \mu$，$\mathbb{D}(X_k) = \sigma^2 > 0$，记$B_n^2 = \sum_{k=1}^{n} \sigma_k^2$ 若存在$\delta > 0$，使得

$$\lim_{n\to\infty} \frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}\left(|X_k - \mu|^{2+\delta}\right) = 0$$

则随机变量之和的标准化变量

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

5 统计

5.1 样本与抽样

定义 19 (简单随机样本). 在相同的条件下对总体X进行n次重复的、独立的观察，将n次观察结果按照实验的次序记为$X_1, X_2, \ldots, X_n$，则这些变量都**相互独立**且与X有**相同分布**

5.2 抽样分布

5.2.1 基本概念

定义 20 (统计量). 样本平均

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

样本方差（期望估计量$\mathbb{E}\left(()\, S^2\right) = \sigma^2$）

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

样本k阶中心矩

$$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$$

经验分布函数

$$F_n(x) = \frac{1}{n}S(x)$$

其中$S(x)$是$X_1, X_2, \ldots, X_n$中不大于x的随机变量的个数

定理 12. 设总体X（不管服从什么分布），均值为μ，方差为σ^2，$X_1, X_2, \ldots, X_n$为来自X的一个样本，$\bar{X}$为样本均值，S^2为样本方差，则

$$\mathbb{E}\left(\bar{X}\right) = \mu, \mathbb{D}\left(\bar{X}\right) = \sigma^2/n$$

定理 13. 设$X_1, X_2, \ldots, X_n$是来自正态总体$N(\mu, \sigma^2)$的样本，则

1. $\bar{X} \sim N(\mu, \sigma^2/n)$（注意$\mathbb{D}\left(X\right) = (\sigma/\sqrt{n})^2$）
2. $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
3. $\bar{X}$与S^2相互独立
4. $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

5.2.2 常见的抽样分布

1. χ^2分布

   设$X_1, X_2, \ldots, X_n$是来自总体$N(0,1)$的样本，则统计量

$$\chi^2 = X_1^2 + X_2 + \cdots + X_n^2 \sim \Gamma\left(\frac{n}{2}, 2\right)$$

服从自由度为n的χ^2分布，记为$\chi^2 \sim \chi^2(n)$

$$
\begin{aligned}
f(x) &= \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1}\mathrm{e}^{-x/2},\, x>0 \\
\mathbb{E}\left(\chi^2\right) &= n \\
\mathbb{D}\left(\chi^2\right) &= 2n
\end{aligned}
$$

χ^2具有可加性，因Gamma分布有可加性

$$
\chi_1^2(n_1) + \chi_2^2(n_2) \sim \chi^2(n_1+n_2)
$$

2. t分布/Student分布
   设$X \sim N(0,1)$，$Y \sim \chi^2(n)$，且X,Y相互独立，则

$$
t = \frac{X}{\sqrt{Y/n}}
$$

   服从自由度为n的t分布，记为$t \sim t(n)$

$$
h(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n}\,\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2},\, t \in (-\infty, +\infty)
$$

3. F分布
   设$U \sim \chi^2(n_1)$，$Y \sim \chi^2(n_2)$，且U,V相互独立，则

$$
F = \frac{U/n_1}{V/n_2}
$$

   服从自由度(n_1,n_2)的F分布，记为$F \sim F(n_1,n_2)$

$$
\psi(x) = \frac{\Gamma[(n_1+n_2)/2](n_1/n_2)^{n_1/2}x^{(n_1/2)-1}}{\Gamma(n_1/2)\Gamma(n_2/2)[1+(n_1x/n_2)]^{(n_1+n_2)/2}},\, x>0
$$

5.3 参数估计

5.3.1 估计量

定义 21 (估计). $X_1, X_2, \ldots, X_n$为独立随机变量，从有参数$\mu, \sigma, \theta, \ldots$的分布$f$中得到，对参数$\theta$的估计是函数$T(X_1, \ldots, X_n)$，称$T$是期望*(expected)*估计，若

$$
\mathbb{E}\left(T(X_1, \ldots, X_n)\right) = \theta
$$

相合*(probable)*的估计，若

$$
\mathbb{P}\left(|T(X_1, \ldots, X_n) - \theta| > \varepsilon\right) \to 0,\, n \to \infty
$$

$$
\begin{aligned}
\mathbb{E}\left(S^2\right) &= \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right) \\
&= \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1^n} X_i^2 - n\bar{X}^2\right) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}\left(X_i^2\right) - n\mathbb{E}\left(\bar{X}^2\right) \qquad \text{由}\sigma^2 = \mathbb{E}\left(X^2\right) - \mu^2 \\
&= \frac{1}{n-1} n(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right) \qquad \text{由}\mathbb{D}\left(\bar{X}\right) = \frac{\sigma^2}{n} = \mathbb{E}\left(\bar{X}^2\right) - \mu^2 \\
&= \sigma^2 = \mathbb{D}\left(X\right)
\end{aligned}
$$

5.3.2 矩估计

设X为随机变量，概率密度为$f(x;\theta_1,\theta_2,\ldots,\theta_n)$，则$X$的前$k$阶矩

$$
\mu_l(\theta_1,\theta_2,\ldots,\theta_k) = \int_{-\infty}^{\infty} x^l f(x;\theta_1,\theta_2,\ldots,\theta_k)\,\mathrm{d}x,\ l = 1,2\ldots,k
$$

k个方程组便可解得k个估计量$\hat{\theta}_l$

5.3.3 最大似然估计法

$$
L(\theta) = L(x_1,x_2,\ldots,x_n;\theta) = \prod_{i=1}^{n} p(x_i;\theta),\ \theta \in \Theta
$$

5.4 可视化

1. 直方图：矩形宽度$\dfrac{f_i}{n}\Big/\Delta$，$\Delta$为组距
2. 箱线图：最小值min，第一四分位数Q_1，中位数M，第三四分位数Q_2，最大值max

$$
q\text{分位数} \quad x_p = \begin{cases} x_{[np]+1} & np \notin \mathbb{Z} \\ \frac{1}{2}[x_{np} + x_{np+1}] & np \in \mathbb{Z} \end{cases}
$$