Concise Notes of Probability Theory

Zhen Hy

2022 年 2 月 3 日

1 重要的知识点

$$\frac{\bar{X}-\mu}{S/\sqrt{n}}, \quad \text{Chebyshev's Inequality}$$

2 Basic Knowledge

Definition 1 (概率, P)**.** 满足：$P(A) > 0$, $P(\Omega) > 0$, $P(\cup_i A_i) = \sum_i P(A_i)$.

Theorem 1 (Bayes)**.**

$$P(B_i|A) = \frac{P(AB_i)}{P(A)} = \frac{P(AB_i)}{\sum_j P(A|B_j)P(B_j)}$$

Definition 2. 两两独立，相互独立

3 Random Variables

Definition 3. 分布函数 F，分布律/列 P（离散），概率密度函数（连续）

Definition 4. *In high dimension:*

- 联合分布函数 $F(x,y) = P(X \le x, Y \le y)$
- 边缘分布函数 $F_X(x), F_Y(y)$
- 条件概率密度 $f_{Y|X} = f(x,y)/f_X(x)$
- 独立 $f(x,y) = f_X(x)f_Y(y)$

Definition 5 (Expectation 期望)**.** $\sum_{k=1}^{+\infty} |x_k| p_k$ 或 $\int_{-\infty}^{+\infty} x f(x) \mathrm{d}x$ 绝对收敛，则该随机变量期望存在。

3.1 Classic Distributions

2-dimentional Normal Distribution $N(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2; \rho)$:

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

Distributions	$0-1$	Bernoulli(二项)	几何	Poisson
记号	$B(1,p)$	$B(n,p)$	$G(p)$	$P(\lambda)$
Domain	0, 1	$1,\ldots,n$	$1,\ldots,+\infty$	$1,\ldots,+\infty$
$\Pr(X=k)$	$p^k(1-p)^{1-k}$	$C_n^k p^k(1-p)^k$	$(1-p)^{k-1}p$	$e^{-\lambda}\ \lambda^k/k!$
$E[X]$	p	np	$1/p$	λ
$D(X)$	$p(1-p)$	$np(1-p)$	$(1-p)/p^2$	λ

表 1: Discrete Distribution

Distributions	Uniform 均匀	Exponential 指数	Normal 正态
记号	$U(a,b)$	$E(\lambda)$	$N(\mu,\sigma^2)$
Domain	(a,b)	$(0,+\infty)$	$\mathbb{R}$
$f_X(x)$	$1/(b-a)$	$\lambda e^{-\lambda x}$	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$
$E[X]$	$(a+b)/2$	$1/\lambda$	μ
$D(X)$	$(b-a)^2/12$	$1/\lambda^2$	σ^2

表 2: Continuous Distribution

3.2 Theorems and Properties

Theorem 2. 卷积: $Z = X + Y \implies f_Z(z) = f_X(x) * f_Y(y)$.可用分布函数直接推倒。

Property 1. *For some distribution: (这些分布是“稳定的”。)*

- *If* $X \sim N(\mu,\sigma^2)$, *then* $\frac{X-\mu}{\sigma} \sim N(0,1)$.
- *If* $(X,Y) \sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho)$, *then* $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2)$
- *If* $X \sim P(\lambda_1)$ *and* $Y \sim P(\lambda_2)$, *then* $X+Y \sim P(\lambda_1+\lambda_2)$.
- *If* $X \sim B(n,p)$ *and* $Y \sim B(m,p)$, *then* $X+Y \sim B(n+m,p)$.

Theorem 3 (* 多维随机变量函数的联合分布). 了解即可（会算线性变换就行？）

$(X,Y) \to f_{(X,Y)}(x,y)$，设 $\begin{cases} u = u(x,y) \\ v = v(x,y) \end{cases}$ 且 $\begin{cases} x = h(u,v) \\ y = s(u,v) \end{cases}$ 同时

$$J(u,v) = \begin{vmatrix} \partial h/\partial u & \partial h/\partial v \\ \partial s/\partial u & \partial s/\partial v \end{vmatrix} \neq 0$$

那么

$$f_{(U,V)}(u,v) = f_{(X,Y)}[h(u,v), s(u,v)] \times |J(u,v)|$$

Property 2 (期望，方差). *May be useful:*

- $E[X+Y] = E[X] + E[Y]$.
- X, Y 独立，$E[XY] = E[X]E[Y]$ *and* $D(X+Y) = D(X) + D(Y)$.
- $D(CX) = C^2 D(X)$.

Definition 6 (协方差,相关系数). $cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)]$ *and* $\rho = \frac{cov(X,Y)}{\sqrt{D(X)D(Y)}} \in [-1,1]$. ⋆ 注：若独立，则一定不相关；反之不然。

4 中心极限定理

Theorem 4 (Markov's Inequality).

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Theorem 5 (Chebyshev's Inequality).

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

or

$$P(|X-\mu| < \varepsilon) > 1 - \frac{\sigma^2}{\varepsilon^2}$$

记：有一对相反的不等号。*Also, Chebyshev's Inequality is* ***significant*** *when prooving "Laws of Large Number".*

Theorem 6 (Laws of Large Number). *Here's some different forms of the law:*

- *Chebyshev.* 两两独立且不相关，方差存在且有共同的上界。

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i]\right|\right)$$

  Note: 不相关条件可以换成 $\lim_{n\to+\infty}\frac{1}{n^2}D(\sum_{k=1}^{n} X_k) = 0$.

- *Khintchine.*

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{N} X_k - \mu\right| \geq \varepsilon\right) = 0$$

Theorem 7 (CLT).

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

5 数理统计

Definition 7. 一些统计量如下：

- 样本均值。$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
- 样本方差。$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
- 样本 k 阶原点矩。$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$.
- 样本 k 阶中心矩。$(CM)_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$.

Definition 8. 分数位：

- 上侧 α 分数位。$P(X > x_\alpha) = \alpha$
- 双侧 α 分数位。$P(|X| > x_{\alpha/2}) = \alpha$

Distributions	χ^2 卡方分布	t 分布	F 分布
记号	$\chi^2(n)$	$t(n)$	$F(m,n)$
意义	$\sum_{i=1}^{n} N^2(0,1)$	$\frac{N(0,1)}{\sqrt{\chi^2(n)/n}}$	$\frac{\chi^2(m)/m}{\chi^2(n)/n}$
E[X]	n	/	/
D[X]	2n	/	/

表 3: 常用统计量的分布

Property 3 (F 分布). $\dfrac{1}{F(m,n)} \sim F(n,m)$ *and* $\dfrac{1}{F_{1-\alpha}(m,n)} = F_\alpha(n,m)$

Theorem 8. *May be **significant**.* $X \sim N(\mu, \sigma^2)$.

- $\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$ *or* $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.
- $\dfrac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \left(\dfrac{X_i - \bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$.
- $\bar{X}$, $\dfrac{(n-1)S^2}{\sigma^2}$ 独立。

$$\Longrightarrow \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1).$$

6 点估计

Definition 9. 矩估计法，最大似然法（有一些特殊的情况如 $\hat{\theta} = \min_i \theta_i$）

Definition 10 (无偏性)**.** $E[\hat{\theta}] = \theta$.

Definition 11 (有效性)**.** *If* $D(\hat{\theta}_1) < D(\hat{\theta}_2)$, *then* $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效。

Theorem 9 (Rao-Cramer 不等式)**.**

$$D(\hat{\theta}) \geq I(\theta) = 1 \Big/ \left\{ nE\left[\left(\frac{\partial \ln f(X;\theta)}{\partial \theta}\right)^2 \right] \right\}$$

于是有“有效估计量”的定义。

Definition 12 (一致估计量)**.** $\lim_{n\to+\infty} P\left(\left|\hat{\theta} - \theta\right| < \varepsilon\right) = 1$.

或者：无偏估计量且 $\lim_{n\to+\infty} D(\hat{\theta}) = 0$

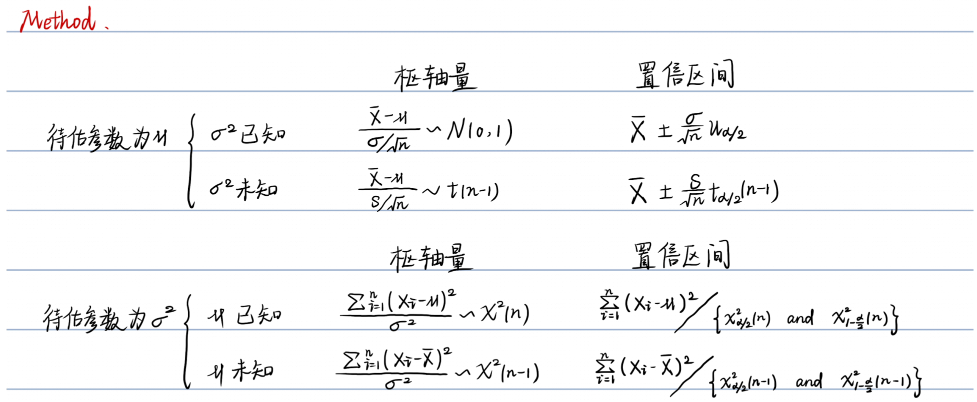

图 1: 单个正态总体参数的置信区间

7 假设检验

Definition 13 (p 值检验法)**.**

一定要注意大于小于号!

过程与置信区间一致，略。