

01 泊松分布

$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!} \Rightarrow X \sim \pi(\lambda)$

对于一个二项分布 $X \sim B(n, p)$, 当 n 很大、 p 很小时, 可以近似为泊松分布, 即 $np \rightarrow \lambda(n \rightarrow +\infty)$.

02 分布函数的性质

- 1. $0 < F(x) \leq 1$
- 2. if $x_1 < x_2, F(x_1) \leq F(x_2)$ (单调不减性)
- 3. $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0, F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$ (归一性)
- 4. $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$ (右连续性)

03 概率密度函数的性质

- 1. $f(x) \geq 0$
- 2. $\int_{-\infty}^{+\infty} f(x) \mathrm{d}x = 1$
- 3. $P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) \mathrm{d}x$
- 4. 若 $f(x)$ 在点 x 处连续, 则有 $F'(x) = f(x)$.

04 指数分布

$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & \text{when } x > 0 \\ 0, & \text{when } x \leq 0 \end{cases} \Rightarrow X \sim E(\theta)$

05 正态分布

$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Rightarrow X \sim N(\mu, \sigma^2)$

标准正态分布函数的性质: $\Phi(-x) = 1 - \Phi(x)$

利用查表法求概率: $P\{c \leq X \leq d\} = \Phi(\frac{d-\mu}{\sigma}) - \Phi(\frac{c-\mu}{\sigma})$

06 求随机变量函数 $Y = g(x)$ 的概率密度

- 分布函数法 (通用)
 - 1. 先求 Y 的分布函数 $F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\}$ (反解出 $P\{X \leq h(y)\} = \int_{-\infty}^{h(y)} f_X(x) \mathrm{d}x$ 的形式)
 - 2. 再求 Y 的概率密度函数 $f_Y(y) = F_Y'(y)$
- 定理法
若 $g(x)$ 处处可导且单调, 则有 $f_Y(y) = f_X(h(y)) |h'(y)|$, when $\alpha < y < \beta$, 其中 $\alpha = \min(g(-\infty), g(+\infty)), \beta = \max(g(-\infty), g(+\infty)), h(y)$ 是 $g(x)$ 的反函数。

07 二维随机变量分布函数的性质

$P(x_1 < X < x_2, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_2, y_1) + F(x_1, y_1) - F(x_1, y_2) \geq 0$

08 二维正态分布

$(X, Y) \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$

09 边缘分布函数

$F_X(x) = F(x, +\infty), f_X(x) = \int_{-\infty}^{+\infty} f(x, y) \mathrm{d}y$

$F_Y(y) = F(+\infty, y), f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) \mathrm{d}x$

10 独立性

X 和 Y 相互独立 \Leftrightarrow

$P\{X = x_i, Y = y_j\} = P\{X = x_i\}P\{Y = y_j\}$ (离散型)

$f(x, y) = f_{X(x)}f_{Y(y)}$ (连续型)

11 条件分布

$P\{X = x_i \mid Y = y_j\} = \frac{P\{X=x_i, Y=y_j\}}{P\{Y=y_j\}}.$

$F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(x|y) \mathrm{d}x = \int_{-\infty}^x \left[\frac{f(x, y)}{f_Y(y)} \right] \mathrm{d}x.$

12 连续型随机变量函数的分布

- $Z = X + Y$ 的分布
 $f_Z(z) = \int_{-\infty}^{+\infty} f(z - y, y) \mathrm{d}y = \int_{-\infty}^{+\infty} f(x, z - x) \mathrm{d}x$
- $Z = \frac{X}{Y}$ 的分布
 $f_Z(z) = \int_{-\infty}^{+\infty} |y| f(yz, y) \mathrm{d}y = \int_{-\infty}^{+\infty} |x| f(x, xz) \mathrm{d}x$
- $M = \max(X, Y)$ 和 $N = \min(X, Y)$ 的分布
 $F_{\max}(z) = F_X(z)F_Y(z)$
 $F_{\min}(z) = 1 - [1 - F_X(z)][1 - F_Y(z)]$

13 连续型数学期望

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

14 方差的定义

$D(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$ 。
根据定义, 方差实际上是随机变量 X 的函数 $g(X) = (X - E(X))^2$ 的数学期望。

15 切比雪夫不等式

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

16 协方差的定义

协方差: $\text{Cov}(X, Y) = E\{(X - E(X))(Y - E(Y))\} = E(XY) - E(X)E(Y)$ 。
相关系数: $\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{D(X) \cdot D(Y)}} \in [-1, 1]$ 。
 $D(X + Y) = D(X) + D(Y) + 2 \text{Cov}(X, Y)$
 $= D(X) + D(Y)$ (特别地, 当 X 和 Y 独立时)

17 数学期望、方差、协方差的性质

$E(CX) = CE(X), E(aX + bY) = aE(X) + bE(Y)$
 $D(CX) = C^2D(X), D(X \pm Y) = D(X) + D(Y), D(X + C) = D(X)$ 。
 $\text{Cov}(X, Y) = \text{Cov}(Y, X), \text{Cov}(X, X) = D(x), \text{Cov}(X, a) = 0, \text{Cov}(aX + bY) = ab \text{Cov}(X, Y), \text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$ 。

18 特殊分布的期望和方差

- 1. 泊松分布 $X \sim P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$: $E(X) = \lambda, D(x) = \lambda$ 。
- 2. 指数分布 $f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$ (when $x > 0$): $E(X) = \theta, D(X) = \theta^2$ 。
- 3. 均匀分布 $f(x) = \frac{1}{b-a}$ (when $a < x < b$): $E(X) = \frac{a+b}{2}, D(X) = \frac{(b-a)^2}{12}$ 。

19 矩

$E(X^k)$: k 阶 (原点) 矩。
 $E(X^k Y^l)$: $k + l$ 阶混合矩。
 $E\{(X - E(X))^k\}$: k 阶中心矩。
 $E\{(X - E(X))^k (Y - E(Y))^l\}$: $k + l$ 阶混合中心矩。

20 大数定理

- 切比雪夫大数定理: 随机变量相互独立, 且具有相同的数学期望和方差: $E(X_k) = \mu, D(X_k) = \sigma^2$, 作前 n 个随机变量的算术平均 $\bar{X} = \frac{\sum_{k=1}^n X_k}{n}$, 则有:
 $\lim_{n \rightarrow \infty} P\{|\bar{X} - \mu| < \varepsilon\} = 1$, 即 \bar{X} 依概率收敛于 μ 。

- 伯努利大数定理： n_A 是n次独立重复试验中事件A发生的次数， p 是事件A在每次试验中发生的概率 $\Rightarrow \frac{n_A}{n}$ 依概率收敛于 μ 。
- 辛钦大数定理：随机变量独立同分布，且具有相同的数学期望 $E(X_k)=\mu\Rightarrow \overline{X}$ 依概率收敛于 μ

21 独立同分布的中心极限定理

随机变量独立同分布，且具有相同的数学期望和方差： $E(X_k)=\mu, D(X_k)=\sigma^2$ ，则随机变量之和的标准化变量： $Y_n=\frac{\sum_{k=1}^n X_k-n\mu}{\sqrt{n}\sigma}$ 的分布函数对于任意 x 满足 $\lim_{n\rightarrow\infty} P\left\{\frac{\sum_{k=1}^n X_k-n\mu}{\sqrt{n}\sigma}\leq x\right\}=\Phi(x)$ ，即收敛于标准正态分布的分布函数。

22 简单随机抽样

$X_1, X_2, ..., X_n$ 是F的一个样本，则它们的联合分布函数为 $F^*(x_1, x_2, ..., x_n) = \prod_{i=1}^n F(x_i)$ ，联合概率密度函数（如果存在）为 $f^*(x_1, x_2, ..., x_n) = \prod_{i=1}^n f(x_i)$ 。

23 常用统计量的定义

- 样本均值： $\overline{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 。

对于正态总体的样本， $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ 。

- 样本方差： $S^2 = \frac{1}{n-1} \sum_{i=1}^n \left(X_i - \overline{X}\right)^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\overline{X}^2\right)$ 。

对于正态总体的样本，有 $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ ， \overline{X} 与 S^2 独立。

对于正态总体的样本， $\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$ 。

24 χ^2 分布

- 定义： $\chi^2 = X_1^2 + X_2^2 + ... + X_k^2 \sim \chi^2(n)$ ，其中 X_i 是来自总体 $N(0, 1)$ 的独立随机变量。
- 期望和方差： $E(X) = n, D(X) = 2n$ 。
- 性质：有 $\chi_1^2 \sim \chi^2(n_1), \chi_2^2 \sim \chi^2(n_2)$ ，则 $\chi_1^2 + \chi_2^2 = \chi^2(n_1 + n_2)$ 。

25 t分布

- 定义：设 $X \sim N(0, 1), Y \sim \chi^2(n)$ ，且 X 和 Y 相互独立，则 $t = \frac{X}{\sqrt{\frac{Y}{n}}} \sim t(n)$ 。
- 性质：图形关于 $t = 0$ 对称，在 $t = 0$ 处取到最大值，在n足够大时近似于标准正态分布。

26 F分布

- 定义：设 $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$ ，则 $F = \frac{\frac{X}{n_1}}{\frac{Y}{n_2}} \sim F(n_1, n_2)$ 。
- 性质：若 $F \sim F(n_1, n_2)$ ，则 $\frac{1}{F} \sim F(n_2, n_1)$ 。

27 点估计的矩估计法

用样本矩的连续函数来估计总体矩的连续函数。

- 总体中只有一个待估参数，用样本的一阶矩来估计总体的一阶矩：
$$\begin{cases} A_1 = \frac{1}{n} \sum_{i=1}^n X_i \\ \mu_1 = E(X) \end{cases} \Rightarrow \mu_1 = A_1 \Rightarrow \text{解出一个未知参数}$$
- 总体中有两个待估参数，用样本的前二阶矩来估计总体的前二阶矩：
$$\begin{cases} A_1 = \frac{1}{n} \sum_{i=1}^n X_i \\ A_2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \\ \mu_1 = E(X) \\ \mu_2 = E(X^2) = D(X) + (E(X))^2 \end{cases} \Rightarrow \begin{cases} \mu_1 = A_1 \\ \mu_2 = A_2 \end{cases} \Rightarrow \text{解出两个未知参数}$$

28 点估计的最大似然估计法

- 写出似然函数：

$L(\theta) = L(x_1, x_2, ..., x_n; \theta)$

$= \prod_{i=1}^n p(x_i; \theta)$ (离散型)

$= \prod_{i=1}^n f(x_i; \theta)$ (连续型)

- 选取使得似然函数 $L(\theta)$ 取得最大值的 $\hat{\theta}$ 作为 θ 的估计值。

可以列出对数似然方程（组） $\frac{d \ln L(\theta)}{d \theta} = 0$ ，解方程得到最大似然估计值。

29 估计量的评价标准

- 无偏性： $E(\hat{\theta}) = \theta$ 。
- 有效性： $D(\hat{\theta}_1) \leq D(\hat{\theta}_2)$ ，则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效。
- 相合性（一致性）：当 $n \rightarrow \infty$ 时， $\hat{\theta}$ 依概率收敛于 θ 。

30 区间估计

- 构造一个关于样本的含 θ 的函数 W ，分布中不含有未知参数。
- 取 W 的分布对应的上 $1 - \frac{\alpha}{2}$ 分位点、上 $\frac{\alpha}{2}$ 分位点，则满足 $P\left\{x_{1-\frac{\alpha}{2}} < W(x) < x_{\frac{\alpha}{2}}\right\} = 1 - \alpha$ 。
- 不等式中仅有唯一参数 θ ，解其取值范围得到置信区间。

针对来自正态总体的样本，可以按照下表直接得到参数的区间估计量：

$X \sim N(\mu, \sigma^2)$			$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ 且相互独立		
待估参数	条件	构造的统计量及其分布	待估参数	条件	构造的统计量及其分布
	σ^2 已知	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$		σ_1^2, σ_2^2 已知	$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1)$
均值 μ			均值差 $\mu_1 - \mu_2$		
	σ^2 未知	$\frac{\bar{X} - \mu}{S_n/\sqrt{n}} \sim t(n-1)$		σ_1^2, σ_2^2 未知但 $\sigma_1^2 = \sigma_2^2$	$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_w \sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2)$ 其中 $S_w^2 = \sqrt{\frac{(n_1-1)S_{n_1}^2 + (n_2-1)S_{n_2}^2}{n_1 + n_2 - 2}}$
方差 σ^2	μ 未知	$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$	方差比 σ_1^2/σ_2^2	μ_1, μ_2 未知	$\frac{S_{n_1}^2/S_{n_2}^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$

31 假设检验

- 提出假设 H_0 和备择假设 H_1 ；
- 构造一个统计量，满足假设 H_0 且不含任何未知参数；
- 给定假设 H_0 成立的概率为 $1 - \alpha$ （显著性水平），这个统计量的取值应当不落入一定的范围 W （拒绝域）；
- 根据样本的观察值，计算统计量的观察值 u_0 ，若 t_0 落入拒绝域，则拒绝假设 H_0 。

针对正态总体的参数假设，可以按照下表进行假设检验：

统计假设		条件	构造的统计量及其分布	拒绝域
H_0	H_1			
$\mu = \mu_0$	$\mu \neq \mu_0$	σ^2 已知	$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$	$ u \geq u_{\alpha/2}$
$\mu \leq \mu_0$	$\mu > \mu_0$			$u \geq u_{\alpha}$
$\mu \geq \mu_0$	$\mu < \mu_0$			$u \leq -u_{\alpha}$
$\mu = \mu_0$	$\mu \neq \mu_0$	σ^2 未知	$T = \frac{\bar{X} - \mu_0}{S_n/\sqrt{n}} \sim t(n-1)$	$ t \geq t_{\alpha/2}(n-1)$
$\mu \leq \mu_0$	$\mu > \mu_0$			$t \geq t_{\alpha}(n-1)$
$\mu \geq \mu_0$	$\mu < \mu_0$			$t \leq -t_{\alpha}(n-1)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	μ 未知	$\chi^2 = \frac{n-1}{\sigma_0^2} S_n^2 \sim \chi^2(n-1)$	$\chi^2 \geq \chi_{\alpha/2}^2(n-1)$ 或 $\chi^2 \leq \chi_{1-\alpha/2}^2(n-1)$
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$			$\chi^2 \geq \chi_{\alpha}^2(n-1)$
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$			$\chi^2 \leq \chi_{1-\alpha}^2(n-1)$

针对正态总体的参数假设，可以按照下表进行假设检验：

统计假设		条件	构造的统计量及其分布	拒绝域
H_0	H_1			
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	σ_1^2, σ_2^2 已知	$U = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1)$	$ u \geq u_{\alpha/2}$
$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$			$u \geq u_{\alpha}$
$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$			$u \leq -u_{\alpha}$
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	σ_1^2, σ_2^2 未知但 $\sigma_1^2 = \sigma_2^2$	$T = \frac{\bar{X} - \bar{Y}}{S_w \sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2)$ 其中 $S_w^2 = \sqrt{\frac{(n_1-1)S_{n_1}^2 + (n_2-1)S_{n_2}^2}{n_1 + n_2 - 2}}$	$ t \geq t_{\alpha/2}(n_1 + n_2 - 2)$
$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$			$t \geq t_{\alpha}(n_1 + n_2 - 2)$
$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$			$t \leq -t_{\alpha}(n_1 + n_2 - 2)$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	μ_1, μ_2 未知	$F = \frac{S_{n_1}^2/S_{n_2}^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$	$F \geq F_{\alpha/2}(n_1-1, n_2-1)$ 或 $F \leq F_{1-\alpha/2}(n_1-1, n_2-1)$
$\sigma_1^2 \leq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$			$F \geq F_{\alpha}(n_1-1, n_2-1)$
$\sigma_1^2 \geq \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$			$F \leq F_{1-\alpha}(n_1-1, n_2-1)$