Probability Theories

随机变量的分布

Poisson Distribution

$$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$$
$$E(X)=Var(x)=\lambda$$

Normal Distribution

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

Exponential Distribution

$$f(x)=\lambda e^{-\lambda x}, x>0$$
$$E(X)=\frac{1}{\lambda}$$
$$Var(X)=\frac{1}{\lambda^2}$$

多维随机变量的分布

二元正态分布

若 $X\sim N(\mu_1,\sigma_1^2), Y\sim N(\mu_2^2,\sigma_2^2), \rho_{XY}=\rho$，则 (X,Y) 服从二元正态分布，记作 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$

- 二元正态分布，独立 $\iff$ 不相关
- 令 $Z=aX+bY+c$ 是 X,Y 的线性组合，则有

  $E(Z)=a\mu_1+b\mu_2+c$，

  $Var(Z)=a^2\sigma_1^2+b^2\sigma_2^2+2ab\rho\sigma_1\sigma_2$

X+Y, Min, Max

- $Z=X+Y: F_Z(z)=P\{X+Y\leq z\}$
- $M=max\{X,Y\}: F_M(t)=P\{max\{X,Y\}\leq t\}=P\{X\leq t, Y\leq t\}=F_X(t)\cdot F_Y(t)$
- $N=min\{X,Y\}: F_N(t)=P\{min\{X,Y\}\leq t\}=P\{(X\leq t)\cup(Y\leq t)\}=1-[1-F_X(t)][1-F_Y(t)]$
- $N=min\{X_1,\ldots,X_n\}: F_N(t)=1-\prod_{i=1}^n[1-F_{X_i}(t)]$

随机变量的数字特征

$Var(aX\pm bY)=a^2Var(X)+b^2Var(Y)\pm 2abCov(X,Y)$

$E(X^2)=\mu^2+\sigma^2 \Leftarrow Var(x)=E(X^2)-E(X)^2$

协方差 $Cov(X,Y)=E[(X-E(X))(Y-E(Y))]=E(XY)-E(X)\cdot E(Y)$ 是双线性的

相关系数

$$\rho_{XY}=\frac{Cov(X,Y)}{SD(x)\cdot SD(Y)}$$

独立一定不相关，相关一定不独立

k 阶原点矩 $\mu_k = E(X^k)$

k 阶中心矩 $\nu_k = E((X - E(X))^k)$

称 x_α 为上α分位数，若$P\{X > x_\alpha\} = \alpha$

大数定律

切比雪夫不等式 $P\{|X-\mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}$

伯努利大数定律

$$\lim_{n\to+\infty} P\{|\frac{n_A}{n} - p| \geq \epsilon\} = 0$$

辛钦大数定律

$$\lim_{n\to+\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \geq \epsilon\} = 0$$

林德伯格-莱维中心极限定理

$X_1, \ldots, X_n \ \ i.i.d.,$

$$\lim_{n\to+\infty} P\{\frac{\sum X_i - E(\sum X_i)}{SD(\sum X_i)} \leq x\}$$
$$= \lim_{n\to+\infty} P\{\frac{\sum X_i - n\mu}{\sqrt{n}\sigma} \leq x\} = \Phi(x)$$

抽样分布

χ^2 分布

设$X_1, \ldots, X_n \sim N(0,1)$, 则称$Y = X_1^2 + \ldots + X_n^2$ 服从自由度为 n 的 χ^2 分布

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

t 分布

设$X \sim N(0,1), Y \sim \chi^2(n)$，则称 $t = \frac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布

$$\lim_{n\to+\infty} f_t(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$
$$t_{1-\alpha}(n) = -t_\alpha(n)$$

F 分布

设 $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$, 称 $F = \frac{U/n_1}{V/n_2}$ 服从第一自由度 n_1，第二自由度 n_2 的 F 分布

正态总体的抽样分布

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

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \tag{2}$$

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1) \tag{3}$$

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1) \tag{4}$$

当 $\sigma_1^2=\sigma_2^2$ 时,

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2) \tag{5}$$

where

$$S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

参数估计

评价准则

若估计量 $\hat{\theta}$ 满足 $E(\hat{\theta})=\theta$，则称 $\hat{\theta}$ 是 θ 的无偏估计量

若 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是 θ 的无偏估计量，$Var(\hat{\theta}_1)\le Var(\hat{\theta}_2)$， 则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效

设均方误差 $Mse(\hat{\theta}) \overset{def}{=} E((\hat{\theta}-\theta)^2)$，若 $Mse(\hat{\theta}_1)\le Mse(\hat{\theta}_2)$，则称在均方误差准则下，$\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更优

若 $\lim_{n\to+\infty} P\{|\hat{\theta}_n-\theta|<\epsilon\}=1$, 则称 $\hat{\theta}_n$ 是 θ 的相合估计量，记作 $\hat{\theta}_n \xrightarrow{P} \theta$

正态总体区间估计

待估量	已知量	枢轴量
μ	σ^2	$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
μ	NA	$t=\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$
σ^2	NA	$\chi^2=\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
$\mu_1-\mu_2$	σ_1^2, σ_2^2 已知	$Z=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$
$\mu_1-\mu_2$	$\sigma_1^2=\sigma_2^2$ 但未知	$t=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$
$\frac{\sigma_1^2}{\sigma_2^2}$	NA	$F=\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$