概率统计第十一讲
正态分布

史灵生　清华数学系

1、本讲提要

1. 正态分布
    - 定义
    - 性质
    - 抽样分布
2. 随机列的收敛性
    - 依概率收敛
    - 按分布收敛

2、正态分布定义

定义

设$\mu \in \mathbb{R}^n$，Σ 阶对称正定矩阵，则服从n维（联合）正态分布$N(\mu, \Sigma)$的随机变量X，其（联合）密度为

$$p(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left\{ -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right\}.$$

称$N(0, I)$为n维标准正态分布。

- 由Σ的正定性，存在可逆方阵A，使$\Sigma = AA^T$。
- 令$Y = A^{-1}(X-\mu)$，则$X = AY + \mu$；
- $|J(y)| = |\det(A)| = \sqrt{|\Sigma|}$。
- $p_Y(y) = p(Ay+\mu)|J(y)| = (2\pi)^{-n/2} e^{-y^T y/2}$。
- $Y \sim N(0, I)$。

3、正态分布特征数

$X=(X_1,X_2,\dots,X_n)^T\sim N(0,I)$

- $X_i\sim N(0,1)$，$i=1,\dots,n$。
- 易见，$p_X(x)=p_{X_1}(x_1)p_{X_2}(x_2)\cdots p_{X_n}(x_n)$，
- 所以，$X_1,X_2,\dots,X_n$相互独立；
- $EX=0$，$Cov(X)=(Cov(X_i,X_j))_{n\times n}=I$。

$X=(X_1,X_2,\dots,X_n)^T\sim N(\mu,\Sigma)$

- $Y=A^{-1}(X-\mu)\sim N(0,I)$，
- $X=AY+\mu$。
- $EX=E(AY+\mu)=\mu$，
- $Cov(X)=Cov(AY)=ACov(Y)A^T=AA^T=\Sigma$。

4、正态分布协方差阵

若$X=(X_1,X_2,...,X_n)^T\sim N(\mu,\Sigma)$，则$Cov(X)=\Sigma$。

- $Y=A^{-1}(X-\mu)\sim N(0,I)$，$X=AY+\mu$。
- $Cov(Y_i,Y_j)=\begin{cases}1, & 若 i=j,\\ 0, & 若 i\neq j.\end{cases}$
- $$\begin{aligned}Cov(X_i,X_j)&=Cov\left(\sum_{k=1}^n a_{ik}Y_k+\mu_i,\sum_{l=1}^n a_{jl}Y_l+\mu_j\right)\\ &=Cov\left(\sum_{k=1}^n a_{ik}Y_k,\sum_{l=1}^n a_{jl}Y_l\right)\\ &=\sum_{k,l}a_{ik}a_{jl}Cov(Y_k,Y_l)=\sum_{k=1}^n a_{ik}a_{jk}。\end{aligned}$$
- 协方差阵$Cov(X)=(Cov(X_i,X_j))_{n\times n}=AA^T=\Sigma$。

5、在正态分布下独立与不相关等价

定理

设$X=(X_1,X_2,...,X_n)^T\sim N(\mu,\Sigma)$，则$X_1,X_2,...,X_n$相互独立
$\Leftrightarrow Cov(X_i,X_j)=0,\ i\neq j$，即$\Sigma$是对角的。

证明：

- 必要性显然，下证充分性。
- 设Σ=diag$(\sigma_1^2,\sigma_2^2,...,\sigma_n^2)$，则$\Sigma^{-1}$=diag$(1/\sigma_1^2,1/\sigma_2^2,...,1/\sigma_n^2)$;
- $p_X(x)=\dfrac{1}{\sqrt{(2\pi)^n|\Sigma|}}\exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$

$$=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_i}e^{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}}=\prod_{i=1}^{n}p_{X_i}(x_i)。$$

6、正态分布的特征函数

定理

设$X \sim N(\mu, \Sigma)$，其中$\mu \in \mathbb{R}^n$，Σ是n阶对称正定实常数矩阵。则X的特征函数为

$$\varphi_X(t) = e^{it^T\mu - \frac{1}{2}t^T\Sigma t}.$$

证明：

- 存在可逆方阵A，使得$\Sigma = AA^T$。
- $Y = A^{-1}(X-\mu) \sim N(0, I)$，$X = AY + \mu$。
- 因为$\varphi_Y(t) = \prod_{k=1}^{n} \varphi_{Y_k}(t_k) = \prod_{k=1}^{n} e^{-t_k^2/2} = e^{-t^Tt/2}$，
- 所以$\varphi_X(t) = e^{it^T\mu}\varphi_Y(A^Tt) = e^{it^T\mu - \frac{1}{2}t^T\Sigma t}$。

7、正态分布的特征函数

定理

设$\mu \in \mathbb{R}^n$，Σ是n阶对称正定实常数矩阵。则存在正态分布$X \sim N(\mu, \Sigma)$，使得

$$\phi : \mathbb{R}^n \to \mathbb{C}, \quad \phi(t) = e^{it^T\mu - \frac{1}{2}t^T\Sigma t}$$

是X的特征函数。

证明：

- 根据线性代数的知识，对n阶对称正定实常数矩阵Σ，存在可逆方阵A，使得$\Sigma = AA^T$。
- 令$Y = (Y_1, Y_2, ..., Y_n)^T \sim N(0, I)$，
- 则令$X = AY + \mu$即可。

8、正态分布定义2

定义

n维随机变量X称为服从正态分布$N(\mu,\Sigma)$，若它的特征函数为

$$\varphi_X(t)=e^{it^T\mu-\frac{1}{2}t^T\Sigma t},\quad \Sigma\text{对称正定。}$$

推论（正态分布在线性变换下的不变性）

若矩阵B行满秩且$X\sim N(\mu,\Sigma)$，则$BX+b\sim N(B\mu+b,B\Sigma B^T)$.

证明：

$$\begin{aligned}\varphi_{BX+b}(t)&=e^{it^Tb}\varphi_X(B^Tt)=e^{it^Tb}e^{it^TB\mu-\frac{1}{2}t^TB\Sigma B^Tt}\\&=e^{it^T(B\mu+b)-\frac{1}{2}t^TB\Sigma B^Tt}\text{。}\end{aligned}$$

9、正态分布判断

定理

n维随机变量X服从正态分布，当且仅当对任意$a \in \mathbb{R}^n \setminus \{\mathbf{0}\}$，$a^T X$为一维正态随机变量。

证明：

- 必要性来自正态分布定义2的推论。
- 充分性：若$a \in \mathbb{R}^n \setminus \{\mathbf{0}\}$且$a^T X$为一维正态随机变量，$EX = \mu$，$\mathrm{Cov}(X) = \Sigma$，
- 则$E(a^T X) = a^T \mu$，$Var(a^T X) = a^T \Sigma a > 0$，$\Sigma$正定。
- $\varphi_{a^T X}(t) = \exp\{i a^T \mu t - \frac{1}{2} a^T \Sigma a t^2\}$。
- $\varphi_X(a) = \varphi_{a^T X}(1) = \exp\{i a^T \mu - \frac{1}{2} a^T \Sigma a\}$。

10、样本均值和方差的分布

定理

设$X = (X_1, \ldots, X_n)^T \sim N(0, I)$，其中$I$为单位矩阵，记均值$\bar{X} = \sum_{i=1}^{n} X_i/n$，样本方差$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$。则

1. $\bar{X} \sim N(0, 1/n)$；
2. $\bar{X}$与S^2相互独立；
3. $(n-1)S^2 \sim \chi^2(n-1)$。

证明：（1）

- $E\bar{X} = 0$，$Var(\bar{X}) = 1/n$；
- $\bar{X} \sim N(0, 1/n)$。

11、样本方差的分布

证明：（2）、（3）

- 取一正交矩阵A使其首行为$\left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$，
- 令$Y = AX$，则Y服从n维正态分布，
- 并且$EY = 0$，$\mathrm{Cov}(Y) = Cov(AX) = AA^T = I$。
- 故$Y = (Y_1, \dots, Y_n)^T \sim N(0, I)$。

$$
\begin{aligned}
(n-1)S^2 &= \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 = X^TX - n\bar{X}^2 \\
&= (A^{-1}Y)^T(A^{-1}Y) - Y_1^2 = Y^TAA^{-1}Y - Y_1^2 \\
&= Y^TY - Y_1^2 = Y_2^2 + \cdots + Y_n^2 \\
&\sim \chi^2(n-1),
\end{aligned}
$$

而且$\bar{X} = Y_1/\sqrt{n}$与$S^2 = (Y_2^2 + \cdots + Y_n^2)/(n-1)$独立。

12、样本均值和方差的分布

定理

设$X=(X_1,\ldots,X_n)^T \sim N(\mu e,\sigma^2 I)$，其中$e$为全一向量，$I$为单位矩阵，记均值$\bar{X}=\sum_{i=1}^{n} X_i/n$，样本方差$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$。则

1. $\bar{X}$与S^2相互独立；
2. $\bar{X}\sim N(\mu,\sigma^2/n)$；
3. $\frac{(n-1)S^2}{\sigma^2}\sim \chi^2(n-1)$；
4. $\frac{\sqrt{n}(\bar{X}-\mu)}{S}\sim t(n-1)$。

13、样本均值和方差的分布

证明：

- 作标准化：$Y=(X-\mu e)/\sigma$，则$Y\sim N(0,I)$。

（2） $\bar{X}=e^TX/n=e^T(\sigma Y+\mu e)/n=\sigma e^TY/n+\mu=\sigma\bar{Y}+\mu$

$\sim N(\mu,\sigma^2/n)$。（因为$\bar{Y}\sim N(0,1/n)$）

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

$$=\sum_{i=1}^n(Y_i-\bar{Y})^2=(n-1)S_Y^2\sim\chi^2(n-1)。$$

（1） 样本均值$\bar{X}=\sigma\bar{Y}+\mu$和样本方差$S^2=\sigma^2S_Y^2$相互独立，

（4） $$\frac{\sqrt{n}(\bar{X}-\mu)}{S}=\frac{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}}\sim t(n-1)。$$

14、样本均值和方差比的分布

推论

设$X=(X_1,\ldots,X_m)^T\sim N(\mu_1 e,\sigma_1^2 I)$，$Y=(Y_1,\ldots,Y_n)^T\sim N(\mu_2 e,\sigma_2^2 I)$独立，其中$e$为全一向量，$I$为单位矩阵，记均值$\bar{X}=\sum\limits_{i=1}^{m}X_i/m$，$\bar{Y}=\sum\limits_{i=1}^{n}Y_i/n$，样本方差

$S_X^2=\dfrac{1}{m-1}\sum\limits_{i=1}^{m}(X_i-\bar{X})^2$，$S_Y^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(Y_i-\bar{Y})^2$。则

1. 样本方差之比$\frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2}\sim F(m-1,n-1)$；
2. 当方差相等$\sigma_1=\sigma_2=\sigma$时，有

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{1}{m}+\frac{1}{n}}\sqrt{\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}}}\sim t(m+n-2).$$

15、样本均值和方差比的分布

证明（2）：

- 当方差相等$\sigma_1=\sigma_2=\sigma$时，有
$$\bar{X}-\bar{Y}\sim N\left(\mu_1-\mu_2,\sigma^2\left(\frac{1}{m}+\frac{1}{n}\right)\right),$$
- $$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim N(0,1),$$
- $$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{1}{m}+\frac{1}{n}}\sqrt{\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}}}=\frac{\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}}}{\sqrt{\left(\frac{(m-1)S_X^2}{\sigma^2}+\frac{(n-1)S_Y^2}{\sigma^2}\right)/(m-1+n-1)}}$$
$\sim t(m+n-2)$。

16、依概率收敛

定义

设$\{X_n\}$为一随机变量序列，X为一随机变量。若对任意的$\epsilon > 0$，有$P(|X_n - X| \geq \epsilon) \to 0$（$n \to \infty$），则称$\{X_n\}$依概率收敛于$X$，记作$X_n \xrightarrow{P} X$。

定理

设$\{X_n\}$、$\{Y_n\}$是两个随机变量序列，a，b是两个常数。如果$X_n \xrightarrow{P} a$，$Y_n \xrightarrow{P} b$，则有

1. $X_n \pm Y_n \xrightarrow{P} a \pm b$;
2. $X_n Y_n \xrightarrow{P} ab$;
3. $X_n / Y_n \xrightarrow{P} a/b$，（$b \neq 0$）。

17、加减法

定理

设$\{X_n\}$、$\{Y_n\}$是两个随机变量序列，a，b是两个常数。

如果$X_n \xrightarrow{P} a$，$Y_n \xrightarrow{P} b$，则

1. $X_n \pm Y_n \xrightarrow{P} a \pm b$。

证明：（1）

- $\{|X_n + Y_n - (a+b)| \geq \epsilon\} \subset \{|X_n - a| \geq \epsilon/2\} \cup \{|Y_n - b| \geq \epsilon/2\}$，
- $0 \leq P(|X_n + Y_n - (a+b)| \geq \epsilon)$

  $\leq P(|X_n - a| \geq \epsilon/2) + P(|Y_n - b| \geq \epsilon/2) \to 0$，$(n \to \infty)$，
- $X_n + Y_n \xrightarrow{P} a + b$。
- 类似可证$X_n - Y_n \xrightarrow{P} a - b$。

18、乘法

证明：（2）

- 若$X_n \xrightarrow{P} 0$，则$X_n^2 \xrightarrow{P} 0$：
  $P(|X_n^2| \geq \epsilon) = P(|X_n| \geq \sqrt{\epsilon}) \to 0$，（$n \to \infty$）。
- 若$X_n \xrightarrow{P} a$，则$cX_n \xrightarrow{P} ca$：若$c \neq 0$，则
  $P(|cX_n - ca| \geq \epsilon) = P(|X_n - a| \geq \epsilon/|c|) \to 0$，（$n \to \infty$）。
- 若$X_n \xrightarrow{P} a$，则$X_n^2 \xrightarrow{P} a^2$：
  $X_n - a \xrightarrow{P} 0$，$(X_n - a)^2 \xrightarrow{P} 0$，$2a(X_n - a) \xrightarrow{P} 0$，
  $X_n^2 - a^2 = (X_n - a)^2 + 2a(X_n - a) \xrightarrow{P} 0$，即$X_n^2 \xrightarrow{P} a^2$。
- 同理$Y_n^2 \xrightarrow{P} b^2$，$(X_n + Y_n)^2 \xrightarrow{P} (a+b)^2$。

$$X_nY_n = [(X_n + Y_n)^2 - X_n^2 - Y_n^2]/2 \xrightarrow{P} [(a+b)^2 - a^2 - b^2]/2 = ab.$$

19、除法

证明：（3）

- 先证：$1/Y_n \xrightarrow{P} 1/b$。
- $$P(|1/Y_n - 1/b| \geq \epsilon) = P[|(Y_n - b)/(bY_n)| \geq \epsilon]$$
$$= P\left(\left|\frac{Y_n - b}{bY_n}\right| \geq \epsilon, |Y_n - b| \geq \epsilon\right)$$
$$+ P\left(\left|\frac{Y_n - b}{b^2 + b(Y_n - b)}\right| \geq \epsilon, |Y_n - b| < \epsilon\right)$$
$$\leq P(|Y_n - b| \geq \epsilon) + P\left(\frac{|Y_n - b|}{b^2 - |b|\epsilon} \geq \epsilon\right)$$
$$= P(|Y_n - b| \geq \epsilon) + P\left[|Y_n - b| \geq \epsilon\left(b^2 - |b|\epsilon\right)\right]$$
$$\to 0, \quad (n \to \infty)。$$
- $X_n/Y_n = X_n \times 1/Y_n \xrightarrow{P} a/b$。

20、按分布收敛

定义

设随机变量X, X_1, X_2，…的分布函数分别为$F(x), F_1(x), F_2(x), \ldots$ 若对$F(x)$的任意连续点x，都有$\lim\limits_{n\to\infty} F_n(x) = F(x)$，则称$\{F_n(x)\}$弱收敛于$F(x)$，记作$F_n(x) \xrightarrow{W} F(x)$；也称$\{X_n\}$按分布收敛于$X$，记作$X_n \xrightarrow{L} X$。

定理

- $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{L} X$;
- 特别地，若$X = c$为常数，则$X_n \xrightarrow{P} c \Leftrightarrow X_n \xrightarrow{L} c$。

- 设X服从$P(X=1) = P(X=-1) = 1/2$，令$X_n = -X$，
- 则X_n与X同分布，故$X_n \xrightarrow{L} X$。
- 但若$0 < \epsilon < 2$，则$P(|X_n - X| \geq \epsilon) = P(2|X| \geq \epsilon) = 1$，
- 即X_n不依概率收敛于X。