续上节，补充几个例子。

例 0.1 设随机变量 $X \sim N(\mu, \Sigma)$，$Y = AX + b$，A 满秩，计算 Y 的分布。

解：变量变换法计算概率密度函数：

$$\begin{aligned} P_Y(y) &= P_X(x)\frac{1}{|det(\frac{\partial y}{\partial x})|} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}exp\{-\frac{1}{2}(x-\mu)^{'}\Sigma^{-1}(x-\mu))\}\frac{1}{||A||} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}exp\{-\frac{1}{2}(y-A\mu-b)^{'}(A^{-1})^{'}\Sigma^{-1}A^{-1}(y-A\mu-b))\}\frac{1}{||A||} \end{aligned}$$

$$\begin{cases} \mu_Y = A\mu + b \\ \Sigma_Y = A\Sigma A^{'} \end{cases}$$

容易验证 $(2\pi)^{\frac{n}{2}}||A|||\Sigma|^{\frac{1}{2}} = (2\pi)^{\frac{n}{2}}|A\Sigma A^{'}|^{\frac{1}{2}}$

例 0.2 设 $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} {\sigma_1}^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & {\sigma_2}^2 \end{pmatrix}\right]$，求 $U = X + Y$。

解：增补变量，$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 1, & 1 \\ 1, & 0 \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}$

$\begin{pmatrix} U \\ V \end{pmatrix}$ 均值为 $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \mu_1 + \mu_2 \\ \mu_1 \end{pmatrix}$

协方差阵为 $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} {\sigma_1}^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & {\sigma_2}^2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$

$= \begin{pmatrix} {\sigma_1}^2 + \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_2 + {\sigma_2}^2 \\ {\sigma_1}^2 & \rho\sigma_1\sigma_2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$

$= \begin{pmatrix} {\sigma_1}^2 + 2\rho\sigma_1\sigma_2 + {\sigma_2}^2 & {\sigma_1}^2 + \rho\sigma_1\sigma_2 \\ {\sigma_1}^2 + \rho\sigma_1\sigma_2 & {\sigma_1}^2 \end{pmatrix}$

故 $U = X + Y \sim N(\mu_1 + \mu_2, {\sigma_1}^2 + 2\rho\sigma_1\sigma_2 + {\sigma_2}^2)$

可见上述补充变量的目的是为了利用保维的结论。

0.1 多维随机变量的特征数

随机变量的分布全面刻画了随机变量的取值和统计规律，而有时我们只需要对随机变量有一个简洁的认识，不求全面，而是从一个侧面描述分布的特征。除了各个分量的均值、方差之外，还有反映多个变量之间关联程度的特征数 (协方差、相关系数)。

0.1.1 期望

定义 0.1 设 $X \triangleq \begin{pmatrix} X_1 \\ X_2 \\ \cdots \\ X_n \end{pmatrix}$，若各分量的期望均存在，则称 $EX \triangleq \begin{pmatrix} EX_1 \\ EX_2 \\ \cdots \\ EX_n \end{pmatrix}$ 为 X 的期望，可以把 E 理解为一个算符。

性质 (1)　随机变量函数的期望：

$$E\left[g(X)\right]=\begin{cases}\sum_k g(X_k)P_X(X_k)\\ \int_{\mathbb{R}^n} g(x)P_X(x)\,dx\end{cases}$$

例 0.3 在长为 a 的棍子上任取两点, 求两点间的平均长度。

解：两点 X 与 Y 均服从 $(0,a)$ 上的均匀分布，且两者独立。$Z=|X-Y|$ 表示两点间的长度。

$$\begin{aligned}E\left[|X-Y|\right]&=\int_{(x,y)\in\mathbb{R}^2}|x-y|P_{XY}(x,y)\,dx\,dy\\&=\int_0^a\int_0^a|x-y|\frac{1}{a^2}\,dx\,dy\end{aligned}$$

性质 (2)　线性。$E\left(aX+bY\right)=aE\left(X\right)+bE\left(Y\right)$。

性质 (3)　若 X 与 Y 独立，则：

$$\begin{aligned}&E\left(XY\right)=E\left(X\right)E\left(Y\right)\\&Var\left(aX+bY\right)=a^2Var\left(X\right)+b^2Var\left(Y\right)\end{aligned}$$

性质 (4)　数学期望代表随机变量取值的平均位置，具体来说，在均方误差意义下，数学期望可视为最小均方误差估计。有如下定理:

定理 0.1 设 $X\in\mathbb{R}^n$，有：

$$E\left[X\right]=\underset{a\in\mathbb{R}^n}{\arg\min}\,E\left[||X-a||^2\right]$$

注:　定理表明，若固定用一个矢量代表随机矢量，期望就是使得均方误差最小的那个代表。说明如下：

$$\begin{aligned}||X-a||^2&=(X-a)^{'}(X-a)\\&=(X-EX+EX-a)^{'}(X-EX+EX-a)\\&=||X-EX||^2+2(X-EX)^{'}(EX-a)+||EX-a||^2\end{aligned}$$

因此，

$$E\left[||X-a||^2\right]=E\left[||X-EX||^2\right]+E\left[||EX-a||^2\right]\geqslant E\left[||X-EX||^2\right]$$

性质 (5)　期望与概率的等价性。

称 I_A 为事件 A 的示性变量，若 $I_A=\begin{cases}1, & \text{如果}A\text{ 发生}\\ 0, & \text{其它}\end{cases}$，则 $E[I_A]=P(A)$。

例 0.4 配对数的均值和方差

N 个人把帽子丢在某房间，将帽子充分混合后让每个人随机地取一顶帽子。求拿到自己帽子的人数的期望值。

解： 设随机变量 $X_k = \begin{cases} 1, & \text{第 } k \text{ 个人恰好拿到自己的帽子} \\ 0, & \text{其它} \end{cases}$ 。

又 X 表示拿到自己帽子的人数，则：X=X_1+X_2+...+X_N

根据古典概型，$E\left[X_k\right] = P\left(X_k = 1\right) = \frac{(N-1)!}{N} = \frac{1}{N}$

故，$E\left[X\right]$=$E\left[X_1\right]$+$E\left[X_2\right]$+...+$E\left[X_N\right] = 1$

延伸： 求解方差 $\mathrm{var}\left[X\right] = E\left[X^2\right] - \left(E\left[X\right]\right)^2$

$E\left[X^2\right]$=$E\left[\left(\sum_{i=1}^{N} X_i\right)\left(\sum_{j=1}^{N} X_j\right)\right] = E\left[\sum_{i=1}^{N}\left(X_i\right)^2 + 2\sum_{1\le i<j\le N}^{N} X_iX_j\right]$

$i<j: E\left[X_iX_j\right] = P\left(X_i = 1\right)P\left(X_j = 1\left|X_i = 1\right.\right) = \frac{1}{N}\frac{1}{N-1}$

总结： 以后涉及到求次数的期望、方差、高阶矩，转化为示性变量和的均值、方差、高阶矩是一个重要思路。

0.1.2 协方差

在多维随机变量中，讨论各分量之间关系，各分量间有时会毫无影响，有时会相互影响。前面介绍了各分量间关系的一个极端情况，就是各分量取值互不影响，相互独立。下面，我们希望用一个特征数对两个变量间相互关联的关系做一个简洁的描述。

定义 0.2 定义 X 与 Y 的协方差 $Cov\left(X,Y\right) \triangleq E\left[\left(X-EX\right)\left(Y-EY\right)\right]$，
特别地，

$$Cov\left(X,X\right) = Var\left(X\right)$$

性质 (1) 线性与对称

$Cov\left(X,a\right) = 0$

$Cov\left(aX+bY,Z\right) = aCov\left(X,Z\right) + bCov\left(Y,Z\right)$

$Cov\left(X,Y\right) = Cov\left(Y,X\right)$

性质 (2) $Cov\left(X,Y\right) = E\left[XY\right] - E\left[X\right]E\left[Y\right]$

证明： $E\left[XY - XE\left(X\right) - E\left(X\right)Y + E\left(X\right)E\left(Y\right)\right] = E\left(XY\right) - E\left(X\right)E\left(Y\right) - E\left(X\right)E\left(Y\right) + E\left(X\right)E\left(Y\right) = E\left(XY\right) - E\left(X\right)E\left(Y\right)$

性质 (3) $Cov\left(X\pm Y\right) = Var\left(X\right) + Var\left(Y\right) \pm 2Cov\left(X,Y\right)$

性质 (4) $Cov\left(X,Y\right)\begin{cases} >0 & X,Y\text{正相关：}X\text{ 与 }Y\text{ 同时增加或同时减少} \\ =0 & X,Y\text{不相关：}X\text{ 增加，}Y\text{ 可增可减；}X\text{ 减少，}Y\text{ 可增可减} \\ <0 & X,Y\text{负相关：}X\text{ 增加而 }Y\text{ 减少，}X\text{ 减少而 }Y\text{ 增加} \end{cases}$

性质 (4) **推论**　X 与 Y 不相关 $\Leftrightarrow E(XY) = E(X)E(Y)$。

例 0.5 相关的正确理解

$X \sim N(\mu, \sigma^2), Y = Z^2$

$E(X \cdot X^2) - E(X)E(X^2) = 0$，但此时 X 与 Y 不独立，显然二者有紧密的依存关系。所以这里的相关应理解为线性相关。

0.1.3　相关系数

定义 0.3 定义 X 与 Y 的相关系数为 $\rho_{XY} = Corr(X,Y) \triangleq \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$

例 0.6 若 $(X,Y) \sim N(\mu_1, \mu_2, {\sigma_1}^2, {\sigma_2}^2, \rho)$，则 $Cov(X,Y) = \rho\sigma_1\sigma_2, Corr(X,Y) = \rho$。

性质　$|\rho_{XY}| \le 1$

注：　其证明可采用施瓦茨不等式，$[Cov(X,Y)]^2 \le Var(X)Var(Y)$。

定理 0.2 $Corr(X,Y) = \pm 1$ 的充分必要条件是 X 与 Y 之间几乎处处有线性关系。

定理 0.3 $\min\limits_{a,b} \quad E\left\{[Y - (aX+b)]^2\right\} = (1 - \rho_{XY}^2)Var(Y)$ 系数最优值，$\hat{a} = \frac{Cov(X,Y)}{Var(X)}, \hat{b} = E(Y) - \hat{a}E(X)$。

定理的说明：　考虑用 X 的线性函数去近似 Y，穷尽所有可能的 X 的线性函数，最小均方误差为 $(1 - \rho_{XY}^2)Var(Y)$。这意味着 $|\rho_{XY}|$ 越大，用 X 的线性函数近似 Y 的程度就越好，X 与 Y 之间的线性相关程度就越强；反之亦然。

极端情况：

1. $|\rho| = 1$, X 与 Y 之间的线性相关程度最强，用 X 线性近似 Y 的均方误差为 0。
2. ρ=0，用 X 线性近似 Y 的均方误差等于 Y 的方差。

0.1.4　协方差阵

协方差和相关系数是对两个变量之间相互关联关系的描述。对一个 n 维随机矢量的分量两两之间的相关关系的描述用协方差阵。

定义 0.4 设 $X \triangleq \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}$，称 $E\left[(X - EX)(X - EX)'\right]$ 为 X 的协方差阵 $Cov(X)$。

将 E 理解为一个算符，$E[(X_i - EX_i)(X_j - EX_j)] = Cov(X_i, X_j)$

$$E\begin{pmatrix} \cdot \\ \vdots \\ \cdot \end{pmatrix} = \begin{pmatrix} E\cdot \\ E\cdot \\ E\cdot \end{pmatrix} \qquad E\begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} = \begin{pmatrix} E\cdot & E\cdot & E\cdot \\ E\cdot & E\cdot & E\cdot \\ E\cdot & E\cdot & E\cdot \end{pmatrix}。$$

性质 (1)　协方差阵对角线上元素为方差，非对角线元素为协方差。

性质 (2) $Cov\left(X\right)$ 为非负定阵。

例 0.7 n 元正态分布，$X \sim N\left(\mu, \Sigma\right), p_X\left(x\right) = \frac{1}{\left(2\pi\right)^{n/2}\left|\Sigma\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(x-\mu\right)^{'}\Sigma^{-1}\left(x-\mu\right)\right\}$

可证：$E\left(X\right) = \mu, Cov\left(X\right) = \Sigma$

同时可以验证：二元正态分布 $N\left(\mu_1, \mu_2, \sigma_1^2, \sigma_1^2, \rho\right)$ 的均值矢量和协方差阵分别为：

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$