Random vectors and estimation theory

$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$, where X_i are RVs / RANDOM VECTOR

$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$

mean $\begin{bmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{bmatrix}$ $\begin{bmatrix} E[Y_1] \\ \vdots \\ E[Y_n] \end{bmatrix}$

Two important matrices associated with RV's

- Correlation matrix : $E[XX^T]$ i,j^{th} entry $E[X_i X_j]$
- Cross-correlation : $E[XY^T]$ i,j^{th} entry $E[X_i Y_j]$
- Covariance matrix : $E[(X-E[X])(X^T-E[X]^T)]$ i,j^{th} entry $\mathrm{Cov}(X_i, X_j)$
- Cross-covariance matrix : $E[(X-E[X])(Y^T-E[Y]^T)]$ i,j^{th} entry $\mathrm{Cov}(X_i, Y_j)$

Claim Correlation and covariance matrices are symmetric, positive semidefinite matrices. If K is positive semidefinite, then there exists a zero-mean random vector X with K as its correlation matrix.

Proof If K is a correlation matrix, then

$$K = E[XX^T] \text{ for some RV } X$$

$$\alpha^T K \alpha = \alpha^T E[XX^T]\alpha = E[(\alpha^T X)(X^T\alpha)] = E[(\alpha^T X)^2] \geq 0$$

α: non-random, $\in \mathbb{R}^n$

For the second claim, suppose that K is positive semidefinite. Then, K has non-negative eigenvalues $\lambda_1, \dots, \lambda_n$. U the orthonormal matrix formed from the eigenvectors.

Let Y_i be independent RVs, s.t. $E[Y_i]=0$, $E[Y_i^2]=\lambda_i$

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

Then, $\mathrm{Cov}(Y,Y) = \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_n \end{bmatrix}$

Define $X = UY$

Then, $E[X] = UE[Y] = 0$ and

$$\mathrm{Cov}(X,X) = \mathrm{Cov}(UY, UY) = E[UYY^TU^T] = UE[YY^T]U^T$$
$$= U\Lambda U^T = K$$

One more definition :

Characteristic function $\phi_X(u) = E[\exp(j u^T X)]$

Orthogonality principle

Let us start with the case of scalar RVs.
Let X be a scalar RV that we cannot observe (directly). We can try to "estimate" X either via repeated experiments char. X, or via the observation of some other RV Y.
Say, have complicated formula, do not want to drag RV in it, prefer to have a const. b be a "proxy" for X

What is the best choice for b?
Intuitively, mean should be a "good" proxy. Formally, if we seek to find const. b that minimizes MSE = mean square error $E[(X-b)^2]$, we get

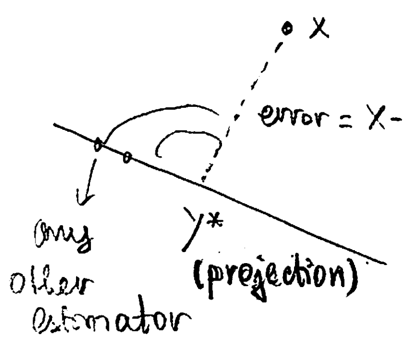
$$E[(X-b)^2] = E[(X-E[X]+E[X]-b)^2] = E[(X-E[X])^2]$$
$$+ 2\underbrace{E[(X-E[X])(E[X]-b)]}_{0} + (E[X]-b)^2 = var(X) + (E[X]-b)^2$$

$b^* = E[X]$
minimizes the error, which in that case equals the variance

Key to the simple proof was

$$E[(X-E[X])] = 0$$

i.e. $E[\underbrace{(X-b^*)}_{\text{error of optimal estimator}} a] = 0 \quad \forall a \in \mathbb{R}$, const. (any estimator)



Formally:
Recall that X, Y are said to be uncorrelated if $E[XY] = E[X]E[Y]$
X, Y are orthogonal if $E[XY] = 0$, $X \perp Y$
Above orthogonality principle asserts that $(X - E[X]) \perp c$

$L^2(\Omega, \mathcal{F}, P)$: set of all RV's on given PS with finite second moment (why finite - will use MSE)

$\mathcal{V} \subseteq L^2(\Omega, \mathcal{F}, P)$ s.t.

a) $\mathcal{V}$ linear class: If $Z_1 \in \mathcal{V}, Z_2 \in \mathcal{V}$, then $a_1 Z_1 + a_2 Z_2 \in \mathcal{V}$, $\forall$ const. a_1, a_2

b) $\mathcal{V}$ is closed in MS sense: $Z_1, Z_2, \ldots \in \mathcal{V}$, $Z_n \xrightarrow{m.s.} Z_\infty \Rightarrow Z_\infty \in \mathcal{V}$

Let $\mathcal{D}$ be as above ②

a) $\exists Z^* \in \mathcal{D}$ such that $E[(X-Z^*)^2] \leq E[(X-Z)^2] \quad \forall Z \in \mathcal{D}$; Z^* is <u>unique</u>

b) $W = Z^*$ iff
$W \in \mathcal{D}$, $\quad (X-W) \perp Z \quad \forall Z \in \mathcal{D}$

c) $\text{MMSE} = E[(X-Z^*)^2] = E[X^2] - E[(Z^*)^2]$

<u>Proof</u> b) Assume that $W \in \mathcal{D}$, $(X-W) \perp Z \quad \forall Z \in \mathcal{D}$

Assume $W \perp$ everything $\Rightarrow$ it will have smallest MSE

$W - Z \in \mathcal{D}$ (linearity) hence $(X-W) \perp (W-Z)$

Then $\underbrace{E[(X-Z)^2]}_{\geq} = E[(X-W+W-Z)^2] = E[(X-W)^2] + E[(W-Z)^2]$
↑ due to orthogonality

i.e. $E[(X-W)^2]$

Assume has smallest MSE show orthogonality

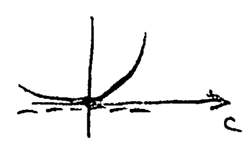
Assume that $W = Z^*$. Clearly, $W \in \mathcal{D}$. Let $Z \in \mathcal{D}$, $c \in \mathbb{R}$, so that $Z^* + cZ \in \mathcal{D}$. Since Z^* is optimal, $E[(X-(Z^*+cZ))^2] \underset{**}{\geq} E[(X-Z^*)^2]$

Expanding

$$E[(X-(Z^*+cZ))^2] = E[(X-Z^*-cZ)^2] = E[(X-Z^*)^2] - E[2(X-Z^*)cZ] + E[c^2Z^2]$$

From $**$

$$\Rightarrow \quad 0 \leq \underbrace{E[c^2Z^2] - 2c\,E[(X-Z^*)Z]}_{f(c)}$$



$f(c) \geq 0$

$$f'(c)\Big|_{c=0} = 0 \quad \Rightarrow$$

$$\left(c^2 E[Z^2] - 2c\,E[(X-Z^*)Z]\right)'\Big|_{c=0} = 0$$

$$\cancel{2}c\,E[Z^2] - \cancel{2}E[(X-Z^*)Z]\Big|_{c=0} = 0$$

$$\Rightarrow E[(X-Z^*)Z] = 0$$

c) $E[(X-Z^*)^2] = E[(X-Z^*)X - (X-Z^*)Z^*] = E[(X-Z^*)X] - \cancelto{0}{E[(X-Z^*)Z^*]}$
$= E[X^2 - Z^*X]$

<u>Easier</u>

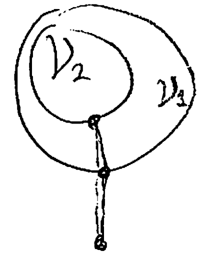
$$E[X^2] = E[(X-Z^*+Z^*)^2] = E[(X-Z^*)^2] + E[Z^{*2}] - \cancelto{0}{2E[(X-Z^*)Z^*]}$$

Summary of properties of projections

Denote the projection of X onto $\mathcal{V}$ as $\Pi_{\mathcal{V}}(X)$

Then

- $\Pi_{\mathcal{V}}(a_1 X_1 + a_2 X_2) = a_1 \Pi_{\mathcal{V}}(X_1) + a_2 \Pi_{\mathcal{V}}(X_2)$

- 

  Both $\mathcal{V}_2, \mathcal{V}_1$ are closed, $\mathcal{V}_1 \subseteq \mathcal{V}_2$

  $$\Pi_{\mathcal{V}_2}(X) = \Pi_{\mathcal{V}_2}(\Pi_{\mathcal{V}_1}(X))$$

  $$E[(X-\Pi_{\mathcal{V}_2}(X))^2] = E[(X-\Pi_{\mathcal{V}_1}(X))^2] + E[(\Pi_{\mathcal{V}_1}(X)-\Pi_{\mathcal{V}_2}(X))^2]$$

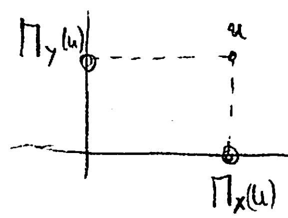
  and $$E[(X-\Pi_{\mathcal{V}_2}(X))^2] \geq E[(X-\Pi_{\mathcal{V}_1}(X))^2]$$

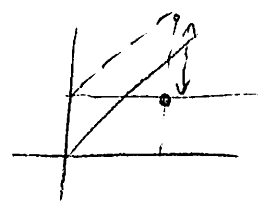
- Both $\mathcal{V}_1, \mathcal{V}_2$ are closed, $\mathcal{V}_1 \perp \mathcal{V}_2$ $\quad \forall Z_1 \in \mathcal{V}_1$, $\forall Z_2 \in \mathcal{V}_2$ $\quad E[Z_1 Z_2] = 0$

  Let $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2 = \{Z_1 + Z_2 : Z_i \in \mathcal{V}_i\}$

  $$\Pi_{\mathcal{V}}(X) = \Pi_{\mathcal{V}_1}(X) + \Pi_{\mathcal{V}_2}(X)$$

  $$E[(X-\Pi_{\mathcal{V}}(X))^2] = E[X^2] - E[(\Pi_{\mathcal{V}_1}(X))^2] - E[(\Pi_{\mathcal{V}_2}(X))^2]$$





The more important case: Estimating X based on observations of another variable Y? Estimator = function $(Y) = f(Y)$

1. All functions f
2. Linear functions f

1 $\quad \mathcal{V} = \{g(Y) : g: \mathbb{R}^m \to \mathbb{R}^n, E[g(Y)^T g(Y)] < \infty\}$ $\qquad$ g has to be Borel measurable

Assume that X, Y have a well-defined joint distribution, and in particular, a joint pdf

$$E[(X-g(Y))^2] = \int_{\mathbb{R}^m} E[(X-g(Y))^2 \mid Y=y] f_Y(y)\,dy$$

where

$$E[(X-g(Y))^2 \mid Y=y] = \int_{\mathbb{R}^m} (x-g(y))^2 f_{X|Y}(x)\,dx$$

Optimal MSE estimator $E[X|Y]$

Orthogonality principle $\quad E[(X-E[X|Y])g(Y)] = 0$

Since $E[\|X-g(Y)\|^2] = \sum_{i=1}^{m} E[(X_i - g_i(Y))^2]$, MMSE estimator analysis carries over from the scalar to the vector case in the obvious way – through componentwise estimation.

$$g^*(Y) = E[X|Y] = \begin{pmatrix} E[X_1|Y] \\ E[X_2|Y] \\ \vdots \\ E[X_m|Y] \end{pmatrix}$$

Linear estimators : $\mathcal{V} = \{c_0 + c_1Y_1 + c_2Y_2 + \dots + c_nY_n : c_0, c_1, \dots, c_n \in \mathbb{R}\}$

$$\hat{X} = AY + b, \quad e = X - AY - b, \quad e_i = (X - AY - b)_i$$

$E[\|X\|^2] < \infty$

$E[\|Y\|^2] < \infty$

$$\Downarrow$$

i) $e_i \perp 1$

ii) $e_i \perp Y_j \quad \forall i, j$

i) $\Rightarrow E[e_i] = 0 \Rightarrow E[X] = AE[Y] + b \Rightarrow b = E[X] - AE[Y]$

ii) $\Rightarrow E[e_i Y_j] = 0$ which is equivalent to $\text{cov}(e_i, Y_j) = 0$, since $E[e_i Y_j] = 0$

$$\text{Cov}(e, Y) = \text{cov}(X - AY - b, Y) = \text{cov}(X, Y) - A\,\text{cov}(Y, Y) - 0 = 0$$

$$\Rightarrow \text{cov}(X, Y) = A\,\text{cov}(Y, Y)$$

$$A = \text{cov}(X, Y)\,\text{cov}^{-1}(Y, Y)$$

if non-singular

Hence, the optimal estimator from the linear class is

$$\hat{E}[X|Y] = E[X] + A(Y - E[Y])$$

$$= E[X] + \text{cov}(X, Y)\,\text{cov}^{-1}(Y, Y)(Y - E[Y])$$

$$\text{Cov}(e) = \text{Cov}(e, e) = \text{Cov}(e, X - g^*(Y)) = \text{Cov}(e, X) - \cancel{\text{Cov}(e, g^*(Y))}$$

$$= \text{cov}(X - E[X] - \text{Cov}(X, Y)\,\text{Cov}(Y, Y)^{-1}(Y - E[Y]), X) = \text{Cov}(X) - \text{Cov}(X, Y)\,\text{Cov}(Y, Y)^{-1}\,\text{Cov}(Y, X)$$

Estimation error:

$$\text{Cov}(e) = \text{Cov}(X) - \text{cov}(X, Y)\,\text{cov}^{-1}(Y, Y)\,\text{cov}(Y, X)$$

Examples

Let $Y = X + Z$ where the signal $X \sim U[-1, 1]$ and noise $Z \sim \mathcal{N}(0, 1)$ are independent.

Find the MMSE for $\text{sign}(X)$: (the sign function)

$$g(y) = E[\text{sign}(X) \mid Y = y] = \int \text{sgn}(x)\, f_{X|Y}(x|y)\, dx$$

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)\, f_X(x)}{f_Y(y)}$$

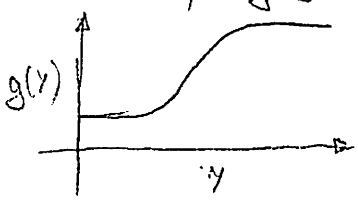
$$f_X(x) = \begin{cases} 1/2, & x \in [-1,1] \\ 0, & \text{otherwise} \end{cases}$$

Since X, Z are independent $\quad Y \mid \{X = x\} \sim N(x, 1)$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{Y|X}(y|x)\, f_X(x)\, dx = \frac{1}{2}\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{|y-x|^2}{2}}\, dx$$

$$= \frac{1}{2}\left(Q(y-1) - Q(y+1) \right)$$

Hence, $$g(y) = \frac{Q(y+1) - 2Q(y) + Q(y-1)}{Q(y-1) - Q(y+1)}$$



sigmoid function

Gaussian RVs

X Gaussian with $\quad f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \; x \in \mathbb{R}$

$\sigma > 0$

$$\Phi_X(u) = \exp\left(-\frac{u^2\sigma^2}{2} + j\mu u \right)$$

Claim $X_1, \dots, X_n$ are independent Gaussian RVs, then any LC $a_1 X_1 + \dots + a_n X_n$ is Gaussian.

Proof suffices to show that claim is true for $n = 2$, $a_1 = a_2 = 1$

$X = X_1 + X_2 \qquad X_1 \sim N(\mu_1, \sigma_1^2), \; X_2 \sim N(\mu_2, \sigma_2^2)$

$\Phi_X(u) = E[e^{juX}] = E[e^{juX_1}]E[e^{juX_2}] = \exp\left(-\frac{u^2\sigma_1^2}{2} + j\mu_1 u\right) \exp\left(-\frac{u^2\sigma_2^2}{2} + j\mu_2 u\right) = \exp\left(-\frac{u^2(\sigma_1^2 + \sigma_2^2)}{2} + ju(\mu_1 + \mu_2) \right)$

Def $\quad (X_i : i \in I)$ is jointly Gaussian if every finite linear combination of variables in the set is Gaussian

$X \sim N(\mu, K) \quad$ mean vector μ, covariance K

Theorem
1) $(X_i : i \in I)$ jointly Gaussian, then each X_i is Gaussian
2) If $(X_i : i \in I)$ are Gaussian, and independent (for any finite collection), then $(X_i : i \in I)$ are jointly Gaussian
3) $(X_i : i \in I)$ jointly Gaussian

   Y_j's finite linear combinations of X_i's ; Z_k's limits of sequences in $(Y_j$ (p., a.s., m.s.)

   $(Y_j : j \in J)$, $(Z_k : k \in K)$ are each jointly Gaussian RVs

! Proposition 2.8 from the lecture notes

X_n RV that is Gaussian, $\forall n$

$X_n \Rightarrow X_\infty$ (a.s., m.s, p, d) $\Rightarrow$ X_∞ is Gaussian !

4) Characteristic function of $X \sim \mathcal{N}(\mu, K)$, $\dim(X) = m$

$$\Phi_X(u) = e^{ju^T\mu - \frac{1}{2}u^T K u}$$

5) If K is diagonal, then components of X are independent

X, Y are jointly Gaussian vectors; they are independent if $Cov(X, Y) = 0$

6) $$f_X(x) = \frac{1}{(2\pi)^{m/2}|K|^{1/2}} \exp\left(-\frac{(x-\mu)^T K^{-1}(x-\mu)}{2}\right)$$

Problem Suppose that X, Y are jointly Gaussian, with parameters $E[X] = \mu_x$, $E[Y] = \mu_y$, $Var(X) = \sigma_x^2$, $Var(Y) = \sigma_y^2$, $E[XY] = \rho_{xy}$. Are the following claims true or false:

1. X, Y are Gaussian. ✓
2. $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ ✓
3. $Z = aX + bY$, $a, b \in \mathbb{R}$, is Gaussian with mean $a\mu_x + b\mu_y$. ✓
4. $\rho_{xy} = 0 \Rightarrow X, Y$ independent ✓
5. $Z = aX + bY$, $a, b \in \mathbb{R}$, is Gaussian with variance $a^2\sigma_x^2 + b^2\sigma_y^2$. ✗

   $Cov(Z, Z) = Cov(aX + bY, aX + bY) = a^2 Var(X) + 2ab\, Cov(X, Y) + b^2 Var(Y)$
6. The linear MMSE of X given Y is

   $$X_L = \mu_x + \rho_{xy}\frac{\sigma_x}{\sigma_y}(Y - \mu_y)$$ ✓

   Linear MMSE = unrestricted MMSE

   $$E[X|Y] = E[X] + Cov(X,Y)\,Cov(Y,Y)^{-1}(Y - E[Y])$$
   $$= \mu_x + \frac{Cov(X,Y)}{\sigma_y^2\sigma_x}\sigma_x(Y - \mu_y) = \mu_x + \rho_{xy}\frac{\sigma_x}{\sigma_y}(Y - \mu_y)$$ ✓

   Recall that $\rho_{xy} = \frac{Cov(X,Y)}{\sqrt{\sigma_x^2\sigma_y^2}} = \frac{Cov(X,Y)}{\sigma_x\sigma_y}$
7. The conditional distribution of $Y|X$ is Gaussian. ✓

   True

Example (Extra)

Let X be an unobservable RV with $E[X]=0$, $var(X)=4$. We observe

$$Y_1 = X + W_1 \qquad E[W_1]=E[W_2]=0$$
$$Y_2 = X + W_2 \qquad Var(W_1)=1, \quad Var(W_2)=4$$

W_1, W_2, X are independent

Find the linear MMSE estimator of X given Y_1, Y_2

$$\hat{X}_L = aY_1 + bY_2 + c$$

Orthogonality principle

1) $E[(X-aY_1-bY_2-c)]=0 \Rightarrow c=0$

2) $E[(X-aY_1-bY_2)Y_1]=0, \quad E[(X-aY_1-bY_2)Y_2]=0$

$\Rightarrow$

$$E[XY_1] - a\,E[Y_1^2] - bE[Y_1Y_2]=0$$
$$E[X(X+W_1)] - a\,E[(X+W_1)^2] - bE[(X+W_1)(X+W_2)]=0$$
$$E[X^2]+\underbrace{E[XW_1]}_{0} - a(E[X^2]+E[W_1^2]) - b(E[X^2]+\underbrace{E[W_1W_2]}_{0})=0$$

$$4-5a-4b=0 \qquad 5a+4b=4$$

Similarly, ... $\qquad 4a+8b=4$

$$\Rightarrow a=\frac{2}{3}, \quad b=\frac{1}{6}$$
$$\hat{X}_L = \frac{2}{3}Y_1 + \frac{1}{6}Y_2$$

Example Suppose X, Y are jointly Gaussian, zero mean, $K=\begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix}$

$E[X^2|Y]=?$

Observe the following: X, Y jointly Gaussian vectors, $y \in \mathbb{R}$

$$X|Y=y \sim \mathcal{N}(X_L(y), Cov(e))$$
$$\text{and} \quad X_L(y) = E[X|Y=y]$$

If $Cov(Y)$ is non-singular

$$E[X|Y=y] = X_L(y) = E[X] + Cov(X,Y)\,Cov(Y)^{-1}(y-E[Y])$$
$$Cov(e) = Cov(X) - Cov(X,Y)\,Cov(Y)^{-1}Cov(Y,X)$$

and if $Cov(e)$ is non-singular

$$f_{X|Y}(x|y) = \frac{1}{(2\pi)^{m/2}|Cov(e)|^{1/2}} \exp\left(-\frac{1}{2}(x-X_L(y))^T Cov^{-1}(e)(x-X_L(y))\right)$$

Proof Consider $X_L(Y)$ = linear function of Y

$e = X - X_L(Y)$ = linear function of X, Y

Orthogonality principle $E[e] = 0$, $Cov(e, Y) = 0$ $\Rightarrow$ Y, e are jointly Gaussian, independent

Next, rewrite $X = e + X_L(Y)$

given $Y = y$, $e \mid Y = y \sim \mathcal{N}(0, Cov(e))$

Hence $X \mid Y = y \sim \mathcal{N}(X_L(y), Cov(e))$

Also, $E[X \mid Y] = X_L(Y)$

Example solution X, Y are jointly Gaussian, zero mean, $K = \begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix}$

Given $Y = y$, the conditional distribution of X is

$$\mathcal{N}\left(\frac{Cov(X,Y)}{Var(Y)} y, \; Cov(X) - \frac{Cov^2(X,Y)}{Var(Y)} \right)$$

$$\parallel$$

$$\mathcal{N}\left(\frac{y}{3}, 3 \right)$$

$$E[X^2 \mid Y = y] = \left(\frac{y}{3}\right)^2 + 3$$

- Suppose X is a RV with $E[X^4] = 30$
  - Derive an upper bound on $P\{|X| \geq 10\}$
  - Find a distribution for X s.t. that the upper bound holds with equality

$$P\{|X| \geq 10\} \leq \frac{E[f(X)]}{f(10)}$$

take $f(X) = X^4$

$$P\{|X| \geq 10\} \leq \frac{E[X^4]}{10^4} = \frac{30}{10^4} = 0.003$$

$$P\{X = 10\} = 0.003$$
$$P\{X = 0\} = 1 - 0.003$$

- Let $U_1, U_2, \ldots$ be a sequence of independent random variables, each uniformly distributed in $[0,1]$

For what values of $c \geq 0$ does there exist a $b > 0$ (depending on c) s.t. $P\{U_1 + \cdots + U_n \geq cn\} \leq e^{-bn} \quad \forall n \geq 1$

$E[U_1] = 1/2 \qquad E[U_1 + \cdots + U_n] = \frac{n}{2}$

want $c\cancel{n} > 1/2\cancel{n} \qquad c > 1/2 \qquad$ If $c = 1/2$, prob is exactly 0.5

$$\%\quad (*)\ P\{U_1 + \cdots + U_n \geq c(U_{n+1} + \cdots + U_{2n})\} \leq e^{-bn} \quad \forall n \geq 1$$

Define $X_i = U_i - c\,U_{n+i}$

X_i's are iid, $E[X_i] = \frac{1}{2} - c\frac{1}{2} = \frac{1-c}{2}$

$(*) = P\{X_1 + \cdots + X_n \geq 0\}$

$\Downarrow$ want $0 > \frac{1-c}{2} \Rightarrow \boxed{c < 1}$