HW1

Ilana Pervoi, Pan Eyal

Problem 1

- **a)** 0
- **b)** No
- **c)** 0

Notes on SPD Matrices, Inner Products, Norms, and Metrics

Problem 1: prove Fact 3 – If A is SPD, then A is invertible, and A^{-1} is SPD too.

Proof: Let A be a SPD matrix. Assume A is not invertible, thus there exist vector $x \neq 0$ such that $Ax = 0$.
$x^T Ax = x^T 0 = 0$ in contradiction to the fact that A is SPD. ($x^T Ax > 0 \ \ \forall x \neq 0$)
Therefore A is invertible, meaning A^{-1} exists.
Let λ be an eigenvalue of A, hence:

$$Av = \lambda v \rightarrow \ A^{-1}Av = A^{-1}\lambda v \rightarrow \ A^{-1}Av = \lambda A^{-1}v \rightarrow Iv = \lambda A^{-1}v \rightarrow \frac{1}{\lambda}v = A^{-1}v$$

According to fact 1, all eigenvalues λ of A are positive. Therefore, all $\frac{1}{\lambda}$ (the eigenvalues of A^{-1}) are positive, and again from fact 1 we obtain that A^{-1} is SPD.

Problem 2: prove Fact 4 – Let Q be an $n \times n$ SPD matrix. Then $\langle\cdot,\cdot\rangle_Q : R^n \times R^n \rightarrow R, \quad \langle\cdot,\cdot\rangle_Q : (x, y) \rightarrow x^T Qy$ is an inner product.

Solution: Let Q be an $n \times n$ SPD matrix and $x, y, z \in R^n, \quad \alpha, \beta \in R$
We will show that the properties of definition 6 hold.

1) $\langle x, y\rangle_Q = x^T Qy \underbrace{=}_{x^T Qy \in R} (x^T Qy)^T = y^T Q^T x \underbrace{=}_{Q\ is\ symmetric} y^T Qx = \langle y, x\rangle_Q$
2) $\langle \alpha x + \beta y, z\rangle_Q = (\alpha x + \beta y)^T Qz = (\alpha x^T + \beta y^T)Qz = \alpha x^T Qz + \beta y^T Qz = \alpha\langle x, z\rangle_Q + \beta\langle y, z\rangle_Q$
3) $\langle x, x\rangle_Q = x^T Qx \underbrace{\geq}_{Q\ is\ SPD} 0$
4) $\Leftarrow$ Let $x = 0_V$, then $x^T Qx = 0_V^T Q 0_V = 0$
   $\Rightarrow$ Let $x^T Qx = 0$, then from SPD definition, $x = 0_V$

Problem 3: prove Fact 8 – Every norm induces a metric: $d(x, y) = ||x - y||$.

Proof: Let $x, y, z \in V(= M)$, we will show that the properties of definition 11 hold.

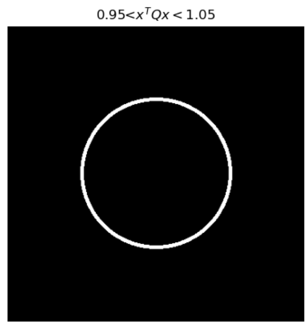
1) $||x - y|| \underbrace{\geqq}_{from\ norm\ definiton} 0$
2) $\Rightarrow$ Let $||x - y|| = 0$ then, from definition, $x - y = 0_V \rightarrow x = y$
   $\Leftarrow$ Let $x = y$, then $x - y = 0_V$ and from definition $||0_V|| = 0$
3) $||x - y|| = ||-1(y - x)|| \underbrace{=}_{from\ norm\ definiton} |-1| \cdot ||y - x|| = ||y - x||$
4) $||x - z|| = ||(x - y) + (y - z)|| \underbrace{\leqq}_{from\ norm\ definiton} ||x - y|| + ||y - z||$

Problem 4: prove Fact 10 – Let Q be an SPD matrix and LL^T denote its Cholesky decomposition.
Then $||x||_Q = \|L^T x\|_{l2}$.

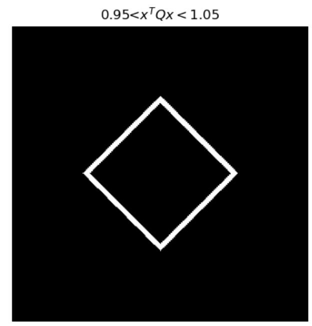
Proof: Let Q be an SPD matrix and LL^T denote its Cholesky decomposition.
$\|x\|_Q = x^T Qx = x^T LL^T x = (L^T x)^T L^T x = \|L^T x\|_{l2}$

Computer Exercise 1:

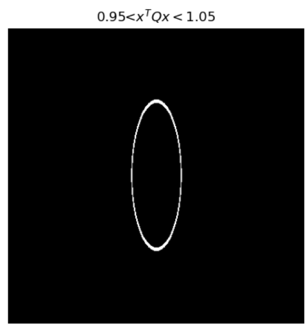
→ The relevant code will be added separately in the submission page.



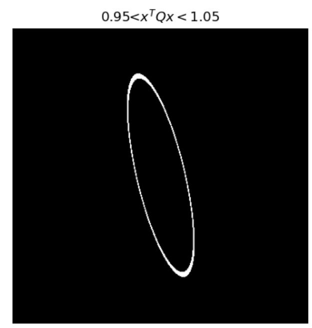
a:



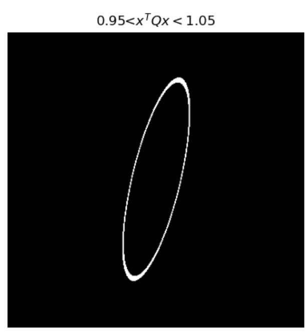
b:



c:



d:



e:

Notes on Convexity

Problem 1:

1) $f: R \to R, f(x) = x^2$ is convex:

   Let assume that $x_1, x_2 \in R$, so:

$$f(tx_1 + (1-t)x_2) \underbrace{\leq}^{?} tf(x_1) + (1-t)f(x_2)$$
$$(tx_1 + (1-t)x_2)^2 \underbrace{\leq}^{?} tx_1^2 + (1-t)x_2^2$$
$$t^2x_1^2 + 2t(1-t)x_1x_2 + (1-t)^2x_2^2 \underbrace{\leq}^{?} tx_1^2 + x_2^2 - tx_2^2$$
$$t^2x_1^2 + 2(t-t^2)x_1x_2 + x_2^2 - 2tx_2^2 + t^2x_2^2 \underbrace{\leq}^{?} tx_1^2 + x_2^2 - tx_2^2$$
$$t^2x_1^2 - tx_1^2 + 2(t-t^2)x_1x_2 + tx_2^2 - 2tx_2^2 + t^2x_2^2 \underbrace{\leq}^{?} 0$$
$$(t^2 - t)x_1^2 + 2(t-t^2)x_1x_2 + (t - 2t + t^2)x_2^2 \underbrace{\leq}^{?} 0$$
$$(t^2 - t)x_1^2 - 2(t^2 - t)x_1x_2 + (t^2 - t)x_2^2 \underbrace{\leq}^{?} 0$$
$$(t^2 - t)(x_1 - x_2)^2 \underbrace{\leq}^{?} 0$$

   This is true because $(x_1 - x_2)^2 \geq 0$ by definition and because $t \in (0,1)$ we will get that $t > t^2$.

2) $f: R \to R, f(x) = |x|$ is convex:

   Let assume that $x_1, x_2 \in R$, so:

$$f(tx_1 + (1-t)x_2) = |tx_1 + (1-t)x_2| =$$
$$= |tx_1 + (1-t)x_2| \underbrace{\leq}_{\substack{by\ the \\ triangle \\ inequality}} |tx_1| + |(1-t)x_2| \underbrace{=}_{since\ t \in (0,1)} t|x_1| + (1-t)|x_2|$$

3) $f: R \to R, f(x) = ax + b$ is convex:

   Let assume that $x_1, x_2 \in R$, so:

$$f(tx_1 + (1-t)x_2) \underbrace{\leq}^{?} tf(x_1) + (1-t)f(x_2)$$
$$a(tx_1 + (1-t)x_2) + b \underbrace{\leq}^{?} t(ax_1 + b) + (1-t)(ax_2 + b)$$
$$atx_1 + a(1-t)x_2 + b \underbrace{\leq}^{?} tax_1 + tb + a(1-t)x_2 + (1-t)b$$
$$b \underbrace{\leq}^{?} tb + (1-t)b$$
$$b \underbrace{\leq}^{?} b\big(t(1-t)\big)$$
$$t - t^2 \underbrace{\geq}_{since\ t \in (0,1)} 0$$

4) $f: R \to R, f(x) = \sin(x)$ is not convex:

   For $x_1 = 0, x_2 = \pi, t = 0.5$:

$$f(tx_1 + (1-t)x_2) = \sin(0.5 * 0 + 0.5 * \pi) = \sin\left(\frac{\pi}{2}\right) = 1$$
$$tf(x_1) + (1-t)f(x_2) = 0.5 * \sin(0) + 0.5 * \sin(\pi) = 0 + 0 = 0$$

   and because $1 \nleq 0$, , $f(x) = \sin(x)$ is not convex.

5) $f: R \to R, f(x) = -x^2$ is not convex:

   For $x_1 = -1, x_2 = 1, t = 0.5$:

$$f(tx_1 + (1-t)x_2) = -(0.5 * -1 + 0.5 * 1)^2 = -(0)^2 = 0$$
$$tf(x_1) + (1-t)f(x_2) = 0.5 * -(-1)^2 + 0.5 * -(1)^2 = -0.5 + -0.5 = -1$$

   and because $0 \nleq -1$, , $f(x) = -x^2$ is not convex.

Notes on Argmin and Argmax

Problem 1: Let $f : S \to R$ be a function from some set S into R. Then $\underset{x\in S}{\operatorname{argmin}} f(x) = \underset{x\in S}{\operatorname{argmax}} -f(x)$.

Proof: Let $f : S \to R$ be a function from some set S into R.

$$\underset{x\in S}{\operatorname{argmax}} -f(x) = \{x: -f(x) \geq -f(x') \ \ \forall x' \in S\} = \{x: f(x) \leq f(x') \ \ \forall x' \in S\} = \underset{x\in S}{\operatorname{argmin}} f(x)$$

Problem 2: For some $g : R \to R$ monotonically non-increasing function, and some $f : S \to R$ it holds that:

$$\underset{x\in S}{\operatorname{argmax}} f(x) \not\subset \underset{x\in S}{\operatorname{argmax}}\, g\big(f(x)\big)$$

Proof: set f as $f(x) = -x^2$ and g as $g(x) = -x$.

It is easy to see that $\underset{x\in R}{\operatorname{argmax}} f(x) = \underset{x\in R}{\operatorname{argmax}}(-x^2) = 0$.

But now: $g\big(f(0)\big) = g(0) = -0 = 0$, while for value $x = 1$, $g\big(f(x)\big) = g(-1) = 1$, and $1 > 0$.

Therefore $\boldsymbol{\underset{x\in R}{\operatorname{argmax}} f(x) = 0 \not\subset \underset{x\in R}{\operatorname{argmax}}\, g\big(f(x)\big)}$.

Problem 3: Let $g : R \to R$ be a monotonically non-decreasing function. Let $f : S \to R$ be a function we seek to maximize. Then, $\underset{x\in S}{\operatorname{argmax}} f(x) \subset \underset{x\in S}{\operatorname{argmax}}\, g\big(f(x)\big)$.

Proof: By the definition of argmax:

$$x_0 \in \underset{x\in S}{\operatorname{argmax}} f(x) \to f(x_0) \geq f(x)\ \forall x \to g\big(f(x_0)\big) \geq g\big(f(x)\big) \forall x \to x_0 \in \underset{x\in S}{\operatorname{argmax}}(g \circ f)(x)$$

Problem 4: Let $r \in R^n$ depend on $\theta \in R^k$. Show that $\underset{\theta\in\mathrm{R^k}}{\operatorname{argmin}} \|r\|_{l2} = \underset{\theta\in\mathrm{R^k}}{\operatorname{argmin}} \|r\|_{l2}^2$.

Proof: By using Fact 5 for $g(x) = x^2$, and $f(x) = \|r(\theta)\|$, we directly deduce that:

$$\boldsymbol{\underset{\theta\in\mathrm{R^k}}{\operatorname{argmin}} \|r\|_{l2} = \underset{\theta\in\mathrm{R^k}}{\operatorname{argmin}} \|r\|_{l2}^2}$$

that is because that $f(x) \geq 0\ \forall x$, and $g(x) = x^2$ is monotonically increasing on $[0, \infty]$.

Problem 5: Let $\sigma \in R_{>0}$ and let $r \in R^n$ depend on $\theta \in R^k$. Show that:

$$\underset{\theta\in\mathrm{R^k}}{\operatorname{argmax}} \frac{1}{(2\pi\sigma)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\frac{\|r\|_{l2}^2}{\sigma^2}\right) = \underset{\theta\in\mathrm{R^k}}{\operatorname{argmin}} \|r\|_{l2}$$

Proof: From Problem 4 we get that:

$$\underset{\theta\in\mathrm{R^k}}{\operatorname{argmin}} \|r\|_{l2} = \underset{\theta\in\mathrm{R^k}}{\operatorname{argmin}} \|r\|_{l2}^2$$

From Fact 1 we get that $\underset{\theta\in\mathrm{R^k}}{\operatorname{argmin}} \|r\|_{l2}^2 = \underset{\theta\in\mathrm{R^k}}{\operatorname{argmax}}(-\|r\|_{l2}^2)$

Now, again from Fact 5, for monotonically increasing function $g(x) = \frac{1}{2}\frac{x}{\sigma^2}$ we get that:

$$\underset{\theta\in\mathrm{R^k}}{\operatorname{argmax}}(-\|r\|_{l2}^2) = \underset{\theta\in\mathrm{R^k}}{\operatorname{argmax}}\left(-\frac{1}{2}\frac{\|r\|_{l2}^2}{\sigma^2}\right)$$

And again for monotonically increasing function $g^*(x) = \exp(x)$ we get that

$$\underset{\theta\in\mathrm{R^k}}{\operatorname{argmax}}\left(-\frac{1}{2}\frac{\|r\|_{l2}^2}{\sigma^2}\right) = \underset{\theta\in\mathrm{R^k}}{\operatorname{argmax}}\left(\exp\left(-\frac{1}{2}\frac{\|r\|_{l2}^2}{\sigma^2}\right)\right)$$

And one last time, for monotonically increasing function $g^{**}(x) = \frac{1}{(2\pi\sigma)^{\frac{n}{2}}}x$ we get that :

$$\underset{\theta\in\mathrm{R^k}}{\operatorname{argmax}}\left(\exp\left(-\frac{1}{2}\frac{\|r\|_{l2}^2}{\sigma^2}\right)\right) = \underset{\theta\in\mathrm{R^k}}{\operatorname{argmax}}\left(\frac{1}{(2\pi\sigma)^{\frac{n}{2}}}\exp\left(-\frac{1}{2}\frac{\|r\|_{l2}^2}{\sigma^2}\right)\right)$$

So overall we achieved that: $\boldsymbol{\underset{\theta\in\mathrm{R^k}}{\operatorname{argmax}} \frac{1}{(2\pi\sigma)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\frac{\|r\|_{l2}^2}{\sigma^2}\right) = \underset{\theta\in\mathrm{R^k}}{\operatorname{argmin}} \|r\|_{l2}}$

__Notes on Linear Least Squares__

Problem 1: Find $\underset{\theta\in\mathrm{R}^k}{\operatorname{argmin}}\|H\theta-y\|_{l2}^2+\lambda_1\theta_1^2+\lambda_3\theta_3^2$

where $H\in R^{N\times 5},\ y\in R^N,\ \theta=[\theta_1\ \theta_2\ \theta_3\ \theta_4\ \theta_5]^T\in R^5, \lambda_1,\lambda_3>0$

__Proof:__ We will bring our problem to a least squares manner:

$$\|H\theta-y\|_{l2}^2+\lambda_1\theta_1^2+\lambda_3\theta_3^2=\left\|\begin{bmatrix} H \\ \begin{bmatrix}\sqrt{\lambda_1} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & \sqrt{\lambda_3} & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}\end{bmatrix}\theta-\begin{bmatrix} y \\ 0_{5\times1}\end{bmatrix}\right\|^2$$

Thus, from linear least squares, the minimizer satisfies the following equation:

$$\begin{bmatrix} H \\ \begin{bmatrix}\sqrt{\lambda_1} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & \sqrt{\lambda_3} & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}\end{bmatrix}^T\begin{bmatrix} H \\ \begin{bmatrix}\sqrt{\lambda_1} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & \sqrt{\lambda_3} & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}\end{bmatrix}\theta=\begin{bmatrix} H \\ \begin{bmatrix}\sqrt{\lambda_1} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & \sqrt{\lambda_3} & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}\end{bmatrix}^T\begin{bmatrix} y \\ 0_{5\times1}\end{bmatrix}$$

$$\rightarrow\left(H^TH+\begin{bmatrix}\lambda_1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & \lambda_3 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}\right)\theta=H^Ty$$

H^TH is always SPD, and the addition of positive values on the diagonal won^' t change that so the new matrix is invertible.

$$\rightarrow\theta=\left(H^TH+\begin{bmatrix}\lambda_1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & \lambda_3 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}\right)^{-1}H^Ty$$

Problem 2: Find $\underset{\hat{x}\in span(v_1,\ldots,v_k)}{\operatorname{argmin}}\|x-\hat{x}\|_{l2}$

__Proof__: as mentioned at the notes, $\hat{x}\in span(v_1,\ldots,v_k)\rightarrow\hat{x}=\sum_{j=1}^k\theta_jv_j$, when $\theta=\begin{bmatrix}\theta_1\\ \vdots\\ \theta_k\end{bmatrix}\in R^k$

and $V=[v_1,\ldots,v_k]\in R^{d\times k}$ and therefore :

$$\hat{x}=V*\theta$$

Meaning the two following problems are equal:

$$\underset{\hat{x}\in span(v_1,\ldots,v_k)}{\operatorname{argmin}}\|x-\hat{x}\|_{l2}\underbrace{=}_{\square^2\ is\ \uparrow}\underset{\hat{x}\in span(v_1,\ldots,v_k)}{\operatorname{argmin}}\|x-\hat{x}\|_{l2}^2=\underset{\theta\in R^k}{\operatorname{argmin}}\|x-V\theta\|_{l2}^2$$

We know how to solve it by LS and the normal equations:

$$\underset{\theta\in R^k}{\operatorname{argmin}}\|x-V\theta\|_{l2}^2=\underset{\theta\in R^k}{\operatorname{argmin}}\|V\theta-x\|_{l2}^2\rightarrow$$

$$V^TV\theta=V^Tx$$

Because V is orthogonal we get:

$$I\theta=V^Tx\rightarrow\theta=V^Tx$$

Notes on Random Vectors

Problem 1: P satisfies finite additivity; namely, if $A1, A2, \ldots, An$ is a finite collection of pairwise disjoint events then $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$.

Proof:

$$\sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} 0 \underbrace{=}_{\substack{by\ infromal\\ definition\ 2\\ (point\ 1)}} \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(\emptyset)$$

Let us denote $A_i = \emptyset$ for every $i > n$, and now:

$$\sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} P(A_i) \underbrace{=}_{\substack{by\ infromal\\ definition\ 2\\ (point\ 3)}} P\left(\bigcup_{i=1}^{\infty} A_i\right)$$

$$= P\left(\bigcup_{i=1}^{n} A_i \cup \bigcup_{i=n+1}^{\infty} A_i\right) \underbrace{=}_{\substack{from\ empty\\ set\ property}} P\left(\bigcup_{i=1}^{n} A_i \cup \emptyset\right) = P\left(\bigcup_{i=1}^{n} A_i\right)$$

Problem 2: Prove that $P(A^c) = 1 - P(A)$

Proof:

$$1 \underbrace{=}_{\substack{by\ infromal\\ definition\ 2\\ (point\ 2)}} P(\Omega) \underbrace{=}_{\substack{definition\ of\\ compliment}} P(A^c \cup A) \underbrace{=}_{\substack{A^c \cap A = \emptyset\\ from\\ fact\ 1}} P(A^c) + P(A)$$

$$\rightarrow\ P(A^c) = 1 - P(A)$$

Problem 3: Let X be a RV whose codomain is R, Find $X^{-1}(B)$ and $P\big(X^{-1}(B)\big)$ where

1) $B = R$

2) $B = \emptyset$

Proof:

1) $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} = \{\omega \in \Omega : X(\omega) \in R\}$

   And because $X: \Omega \rightarrow R$, for every $\omega \in \Omega$, $X(\omega) \in R$, and therefore:

   $$\{\omega \in \Omega : X(\omega) \in R\} = \Omega$$

   And $P\big(X^{-1}(B)\big) = P(\Omega) \underbrace{=}_{\substack{by\ infromal\\ definition\ 2\\ (point\ 2)}} 1$

2) $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} = \{\omega \in \Omega : X(\omega) \in \emptyset\}$

   And because $X: \Omega \rightarrow R$, there is no $\omega \in \Omega$ that $X(\omega) \in \emptyset$, and therefore:

   $$\{\omega \in \Omega : X(\omega) \in R\} = \emptyset$$

   And $P\big(X^{-1}(B)\big) = P(\emptyset) \underbrace{=}_{\substack{by\ infromal\\ definition\ 2\\ (point\ 1)}} 0$

Problem 4: Is the following function, $F: R \rightarrow [0,1]$, a CDF of some RV? $F(x) = \begin{cases} 0 & x \le 42 \\ 1 & x > 42 \end{cases}$

Proof:

No, we will show that $F(x)$ is not right-continuous.

A function is right-continuous at C if $\lim_{x \to c^+} F(x) = F(c)$

Therefore, $\lim_{x \to 42^+} F(x) = \lim_{x \to 42^+} \begin{cases} 0 & x \le 42 \\ 1 & x > 42 \end{cases} = 1 \ne 0 = F(42)$

Means $F(x)$ does not holds right-continuous property of CDF $\rightarrow$ F is not CDF.

Problem 5: Let X be a continuous n-dimensional RV.

1) Let $x \in R^n$. Find $P(X = x) = P\left(X^{-1}(x)\right)$.
2) Give an example for $B \subseteq R^n$ such that B contains at least one element of R^n and $P(X^{-1}(B)) = 0$.
3) Let $\{Bi\}_{i=1}^{\infty}$ be a countable collection of nonempty pairwise disjoint subsets of R^n(i.e., $B_i \neq \emptyset$ for every i and $B_i \cap B_j = \emptyset$ whenever $i \neq j$). Give an example for such a collection where, in addition,

   $P(X \in \bigcup_{i=1}^{\infty} B_i) = P\left(X^{-1}(\bigcup_{i=1}^{\infty} B_i)\right)$ Is zero.

Proof:

1) $P(X = x) = P(x \le X \le x)$

   Let us denote $x = [x_1 \dots x_n]$, $X = [X_1 \dots X_n]$ now:

$$P(x \le X \le x) = P(x_1 \le X_1 \le x_1, \dots, x_n \le X_n \le x_n) = \lim_{\forall i: x_i' \to x_i^-}\left(P(x_1' < X_1 \le x_1, \dots, x_n' < X_n \le x_n)\right)$$

$$= \lim_{\forall i: x_i' \to x_i^-}\left(P(X_1 \le x_1, \dots, X_n \le x_n) - P(X_1 \le x_1', \dots, X_n \le x_n')\right)$$

$$= P(X_1 \le x_1, \dots, X_n \le x_n) - \lim_{\forall i: x_i' \to x_i^-}\left(P(X_1 \le x_1', \dots, X_n \le x_n')\right)$$

$$= P(X \le x) - \lim_{x' \to x^-}\left(P(X \le x')\right) = F(x) - \lim_{x' \to x^-} F(x')$$

   Because X is continuous RV then the CDF $F(x)$ is continuous (definition 9). Thus, $\lim_{x' \to x^-} F(x') = F(x)$

   Now, $F(x) - \lim_{x' \to x^-} F(x') = F(x) - F(x) = 0$.

2) For $B = \{[1 \dots n]\} \in R^n$, denote $x = [1 \dots n]$, therefore: $P\left(X^{-1}(B)\right) = P\left(X^{-1}(x)\right) = P(X = x) \underbrace{=}_{\substack{from \\ section\ 1}} 0$

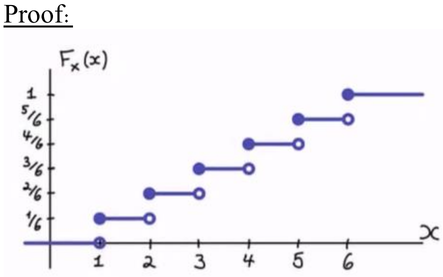
3) For $B_i = \left\{\underbrace{[i \dots i]}_{n\ entries}\right\} \subset R^n$, denote $x_i = \underbrace{[i \dots i]}_{n\ entries}$, therefore:

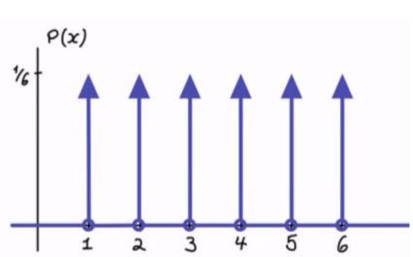
$$(*)\ \ P\left(X^{-1}(B_i)\right) = P\left(X^{-1}(x_i)\right) = P(X = x_i) \underbrace{=}_{\substack{from \\ section\ 1}} 0$$

   We can notice that for every $i \neq j$: $B_i \neq \emptyset$ and $B_i \cap B_j = \emptyset$. Now, for: $P(X \in \bigcup_{i=1}^{\infty} B_i)$ there exist k such that $X \in B_k$, and therefore $X = x_k = [k, \dots, k]$, so: $P(X \in \bigcup_{i=1}^{\infty} B_i) = P(X = x_k) \underbrace{=}_{from\ (*)} 0$

Problem 6: Let $X: \Omega \to R$ stand for the result of rolling a fair die. Thus, for $i = 1, \dots, 6$, $a_i = i$ and $p_i = \frac{1}{6}$ and Draw F(X) and $p(X)$.

Proof:





Problem 7: Write a similar expression for $p(x_3, x_4, x_5)$.

Proof:

$$p(x_3, x_4, x_5) = \sum_{x_1, x_2} p(x) = \sum_{x_1}\sum_{x_2} p(\boldsymbol{x}) = \sum_{x_1}\left(p(x_1, -1, x_3, x_4, x_5) + p(x_1, 1, x_3, x_4, x_5)\right)$$

$$= p(-1, -1, x_3, x_4, x_5) + p(-1,1, x_3, x_4, x_5) + p(1, -1, x_3, x_4, x_5) + p(1,1, x_3, x_4, x_5)$$

Problem 8: Let X stand for the result of rolling a fair die (i.e., X is distributed uniformly on $\{1,2,3,4,5,6\}$. Show that $E(X) = 3.5$.

Proof:

$$\sum_x xp(x) = 1*\frac{1}{6}+2*\frac{1}{6}+3*\frac{1}{6}+4*\frac{1}{6}+5*\frac{1}{6}+6*\frac{1}{6} = 3.5$$

Problem 9: let $x \in R^n$. Then the $n \times n$ matrix xx^T is symmetric; namely, $xx^T = (xx^T)^T$.

Proof: $(xx^T)^T = (\boldsymbol{x}^T)^T x^T = xx^T$

Problem 10: Show that $E(X^T) = \left(E(X)\right)^T$.

Proof:

$$\left(E(X)\right)^T = \left(E\left(\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}\right)\right)^T = \left(\begin{bmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{bmatrix}\right)^T = [E(X_1) \dots E(X_n)] = E([X_1 \dots X_n]) = E\left(\left(\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}\right)^T\right) = E(X^T)$$

Problem 11: Show that $(R_x)_{i,j}$ (i.e., the element in the $i-th$ row and $j-th$ column of R_x)
is $E(X_iX_j)$

Proof:

Denote $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \in R^n$ RV (discrete or continuous – we will see it results in the same ☺)

$(*)$ $g: R^n \to R^{n\times n}$ such that $g(X) = XX^T$

$$R_x = E(XX^T) \underbrace{=}_{(*)} E\left(g(X)\right) = E\left(\begin{bmatrix} X_1^2 & \dots & X_1X_n \\ \vdots & \ddots & \vdots \\ X_nX_1 & \dots & X_n^2 \end{bmatrix}\right) \underbrace{=}_{\substack{Definition\ 23,24- \\ depends\ on\ the\ RV}} \begin{bmatrix} E(X_1^2) & \dots & E(X_1X_n) \\ \vdots & \ddots & \vdots \\ E(X_nX_1) & \dots & E(X_n^2) \end{bmatrix}$$

$$\to (R_x)_{i,j} = E\left(X_iX_j\right)$$

Problem 12: Show that R_x is symmetric, i.e., show that $R_x = R_x^T$.

Proof: We will show that $E\left(X_iX_j\right) = E(X_jX_i)$

For $X_i, X_j \in R^n$ discrete RVs:

$$E\left(X_iX_j\right) = \sum_{x_i}\sum_{x_j} x_ix_jp\left(x_i,x_j\right) = \sum_{x_j}\sum_{x_i} x_jx_ip\left(x_j,x_i\right) = E\left(X_jX_i\right)$$

For $X_i, X_j \in R^n$ continuous RVs:

$$E\left(X_iX_j\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_ix_jp\left(x_i,x_j\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_jx_ip\left(x_j,x_i\right) = E\left(X_jX_i\right)$$

Now, from as we saw in problem 11:

$$R_x = \begin{bmatrix} E(X_1^2) & \dots & E(X_1X_n) \\ \vdots & \ddots & \vdots \\ E(X_nX_1) & \dots & E(X_n^2) \end{bmatrix} \underbrace{=}_{E(X_iX_j)=E(X_jX_i)} \begin{bmatrix} E(X_1^2) & \dots & E(X_nX_1) \\ \vdots & \ddots & \vdots \\ E(X_1X_n) & \dots & E(X_n^2) \end{bmatrix} = R_x^T$$

Problem 13: Show that $(\Sigma_x)_{i,j}$ (i.e., the element in the $i-th$ row and $j-th$ column of Σ_x) is

$$E\left(\left(X_i - E(X_i)\right)\left(X_j - E(X_j)\right)\right)$$

Proof:

Similar to problem 11, denote $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \in R^n$ RV.

$(*)$ $g: R^n \to R^{n\times n}$ such that $g(X) = \left(X - E(X)\right)\left(X - E(X)\right)^T$

$$\Sigma_x = E\left(\left(X - E(X)\right)\left(X - E(X)\right)^T\right) \underbrace{=}_{(*)} E\left(g(X)\right)$$

$$= E\left(\begin{bmatrix} \left(X_1 - E(X_1)\right)^2 & \dots & (X_1 - E(X_1))(X_n - E(X_n)) \\ \vdots & \ddots & \vdots \\ \left(X_n - E(X_n)\right)\left(X_1 - E(X_n)\right) & \dots & \left(X_n - E(X_n)\right)^2 \end{bmatrix}\right)$$

$$\underbrace{=}_{\substack{Definition\ 23,24- \\ depends\ on\ the\ RV}} \begin{bmatrix} E\left(\left(X_1 - E(X_1)\right)^2\right) & \dots & E\left((X_1 - E(X_1))(X_n - E(X_n))\right) \\ \vdots & \ddots & \vdots \\ E\left(\left(X_n - E(X_n)\right)\left(X_1 - E(X_n)\right)\right) & \dots & E\left(\left(X_n - E(X_n)\right)^2\right) \end{bmatrix}$$

$$\to (\Sigma_x)_{i,j} = E\left(\left(X_i - E(X_i)\right)\left(X_j - E(X_j)\right)\right)$$

Problem 14: Show that $(\Sigma_x)_{i,i}$ is the variance of X_i.

Proof: From definition, $V(X_i) = E\left(\left(X_i - E(X_i)\right)^2\right)$

From problem 13, we obtain:

$$(\Sigma_x)_{i,i} = E\left(\left(X_i - E(X_i)\right)\left(X_i - E(X_i)\right)\right) = E\left(\left(X_i - E(X_i)\right)^2\right) = V(X_i)$$

Problem 15: Show that Σ_x is symmetric, i.e., show that $\Sigma_x = \Sigma_x^T$.

Proof: We will show that $E\left((X_i - E(X_i)(X_j - E(X_j)\right) = E((X_j - E(X_j)(X_i - E(X_i))$

Denote $X_i - E(X_i) = Y_i$, $X_j - E(X_j) = Y_j$, hence we need to prove:

$$E\left(Y_i Y_j\right) = E(Y_j Y_i)$$

But we already proved it in problem 12 for discrete and continuous RVs

$$\to (*)\ E\left((X_i - E(X_i)(X_j - E(X_j)\right) = E((X_j - E(X_j)(X_i - E(X_i))$$

Now from problem 13:

$$\Sigma_x = \begin{bmatrix} E\left(\left(X_1 - E(X_1)\right)^2\right) & \dots & E\left((X_1 - E(X_1))(X_n - E(X_n))\right) \\ \vdots & \ddots & \vdots \\ E\left(\left(X_n - E(X_n)\right)\left(X_1 - E(X_n)\right)\right) & \dots & E\left(\left(X_n - E(X_n)\right)^2\right) \end{bmatrix} \underbrace{=}_{(*)}$$

$$\begin{bmatrix} E\left(\left(X_1 - E(X_1)\right)^2\right) & \dots & E\left(\left(X_n - E(X_n)\right)\left(X_1 - E(X_n)\right)\right) \\ \vdots & \ddots & \vdots \\ E\left((X_1 - E(X_1))(X_n - E(X_n))\right) & \dots & E\left(\left(X_n - E(X_n)\right)^2\right) \end{bmatrix} = \Sigma_x^T$$

Problem 16: Prove:

1. $\Sigma = E(XX^T) - \mu\mu^T$
2. $\Sigma = E((X-\mu)X^T)$
3. $\Sigma = E(X(X-\mu)^T)$

Proof:

1. $$\Sigma \underbrace{=}_{Definition} E\left(\big(X-E(X)\big)\big(X-E(X)\big)^T\right) \underbrace{=}_{E(X)=\mu} E\big((X-\mu)(X-\mu)^T\big) = E\big((X-\mu)(X^T-\mu^T)\big)$$
$$= E(XX^T - X\mu^T - \mu X^T - \mu\mu^T) \underbrace{=}_{\substack{Liniarity \\ of\ Expectation}}$$
$$E(XX^T) - E(X\mu^T) - E(\mu X^T) + E(\mu\mu^T) \underbrace{=}_{\mu,\mu^T\ are\ constant\ vectors}$$
$$E(XX^T) - \mu^T E(X) - \mu E(X^T) + \mu\mu^T \underbrace{=}_{\substack{E(X)=\mu \\ E(X^T)=E(X)^T}} E(XX^T) - \mu^T\mu - \mu E(X)^T + \mu\mu^T$$
$$= E(XX^T) - \mu^T\mu - \mu\mu^T + \mu\mu^T = E(XX^T) - \mu^T\mu$$

2. $$\Sigma \underbrace{=}_{Definition} E\left(\big(X-E(X)\big)\big(X-E(X)\big)^T\right) \underbrace{=}_{E(X)=\mu} E\big((X-\mu)(X-\mu)^T\big) = E\big((X-\mu)(X^T-\mu^T)\big) =$$
$$E\big((X-\mu)X^T - (X-\mu)\mu^T\big) \underbrace{=}_{\substack{Liniarity \\ of\ Expectation}} E\big((X-\mu)X^T\big) - E\big((X-\mu)\mu^T\big)$$
$$= E\big((X-\mu)X^T\big) - \mu^T E(X-\mu) \underbrace{=}_{\substack{Liniarity \\ of\ Expectation}} E\big((X-\mu)X^T\big) - \mu^T\big(E(X) - E(\mu)\big)$$
$$= E\big((X-\mu)X^T\big) - \mu^T(\mu-\mu) = E\big((X-\mu)X^T\big)$$

3. $$\Sigma \underbrace{=}_{Definition} E\left(\big(X-E(X)\big)\big(X-E(X)\big)^T\right) \underbrace{=}_{E(X)=\mu} E\big((X-\mu)(X-\mu)^T\big) = E\big((X-\mu)(X^T-\mu^T)\big) =$$
$$E\big(X(X^T-\mu^T) - \mu(X^T-\mu^T)\big) \underbrace{=}_{\substack{Liniarity \\ of\ Expectation}} E\big(X(X^T-\mu^T)\big) - E\big(\mu(X^T-\mu^T)\big)$$
$$\underbrace{=}_{\mu\ is\ a\ constant\ vector} E\big(X(X^T-\mu^T)\big) - \mu E(X^T-\mu^T) = E\big(X(X^T-\mu^T)\big) - \mu\big(E(X^T)-\mu^T\big) \underbrace{=}_{E(X^T)=E(X)^T}$$
$$E\big(X(X^T-\mu^T)\big) - \mu(E(X)^T-\mu^T) = E\big(X(X^T-\mu^T)\big) - \mu(\mu^T-\mu^T) = E\big(X(X^T-\mu^T)\big)$$

Problem 17: Prove:

1. $\mathrm{R}_{\mathrm{YX}} = R_{XY}^T$
2. $\Sigma_{YX} = \Sigma_{XY}^T$
3. $E\left(\big(X-E(X)\big)\big(Y-E(Y)\big)^T\right) = R_{XY} - \mu_X\mu_Y^T$

Proof:

1. $R_{XY}^T = E(XY^T)^T = E((XY^T)^T) = E(YX^T) = R_{YX}$
2. $$\Sigma_{XY}^T = E\left(\big(X-E(X)\big)\big(Y-E(Y)\big)^T\right)^T = E\left(\left(\big(X-E(X)\big)\big(Y-E(Y)\big)^T\right)^T\right)$$
$$= E\left(\big(Y-E(Y)\big)\big(X-E(X)\big)^T\right) = \Sigma_{YX}$$
3. $$E\left(\big(X-E(X)\big)\big(Y-E(Y)\big)^T\right) = E\big((X-\mu_X)(Y-\mu_Y)^T\big) = E\big((X-\mu_X)(Y^T-\mu_Y^T)\big)$$
$$= E(XY^T - X\mu_Y^T - \mu_x Y^T + \mu_x\mu_Y^T) = E(XY^T) - E(X\mu_Y^T) - E(\mu_x Y^T) + E(\mu_x\mu_Y^T)$$
$$= R_{XY} - \mu_x\mu_Y^T - \mu_x\mu_Y^T + \mu_x\mu_Y^T = R_{XY} - \mu_x\mu_Y^T$$

Problem 18: Let $Y = AX + b$ affine transformation. Prove:

1. $\mu_Y = A\mu_X + b$
2. $\Sigma_Y = A\Sigma_X A^T$

Proof:

1. $\mu_Y = E(AX+b) = E(AX) + E(b) \underbrace{=}_{A,b\ are\ constants} AE(X) + b = A\mu_x + b$
2. $\Sigma_Y = E\big((Y-\mu_Y)(Y-\mu_Y)^T\big) = E\big((AX+b-\mu_Y)(AX+b-\mu_Y)^T\big) \underbrace{=}_{1.}$

$$E\big((AX+b-A\mu_x-b)(AX+b-A\mu_x-b)^T\big) = E\left(A(X-\mu_x)\big(A(X-\mu_x)\big)^T\right)$$

$$= E(A(X-\mu_x)(X-\mu_x)^T A^T) \underbrace{=}_{A\ is\ a\ constant} AE\big((X-\mu_x)(X-\mu_x)^T\big)A^T = A\Sigma_X A^T$$

Problem 19: Let $X = [X_1\ \dots\ X_n]^T$ be an n-dimensional RV with mean μ and covariance $\Sigma = \sigma^2 I$.
Find: $V(\sum_{i=1}^n X_i)$

Proof:
Denote: $A = 1^T, b = 0$
$\rightarrow Y = AX + b = 1^T X$ the affine transformation of X.

From fact 10:

$$\Sigma_Y = A\Sigma_X A^T = 1^T\sigma^2 I(1^T)^T = 1^T\sigma^2 I 1 = 1^T\sigma^2 1 = \sigma^2 1^T 1$$
$$= \sigma^2\sum_{i=1}^n 1 = \boldsymbol{\sigma^2 n} \in R^{1\times 1} = V(Y) = V(1^T X) = V\left(\sum_{i=1}^n X_i\right)$$

Problem 20: $A \perp\!\!\!\perp B|C \leftrightarrow p(A|B,C) = p(A|C) \leftrightarrow p(B|A,C) = p(B|C)$

Proof:

We'll prove:

1) $A \perp\!\!\!\perp B|C \leftrightarrow p(A|B,C) = p(A|C)$
2) $A \perp\!\!\!\perp B|C \leftrightarrow p(B|A,C) = p(B|C)$

1) $A \perp\!\!\!\perp B|C \underbrace{\leftrightarrow}_{definition\ 31} P(A\cap B|C) = P(A|C)P(B|C) \underbrace{\leftrightarrow}_{\substack{Bayes\ Theorem\ on\\ P(A\cap B|C),\ P(B|C)}} \frac{P(A\cap B\cap C)}{P(C)} = P(A|C)\frac{P(B\cap C)}{P(C)}$

$$\underbrace{\leftrightarrow}_{multiply\ by\ P(C)} P(A\cap B\cap C) = P(A|C)P(B\cap C)$$

$$\underbrace{\leftrightarrow}_{divide\ by\ P(B\cap C)>0} \frac{P(A\cap B\cap C)}{P(B\cap C)} = P(A|C) \underbrace{\leftrightarrow}_{\substack{Bayes\ Theorem\ on\\ P(A|B,C)}} P(A|B,C) = P(A|C)$$

2) $A \perp\!\!\!\perp B|C \underbrace{\leftrightarrow}_{definition\ 31} P(A\cap B|C) = P(A|C)P(B|C) \underbrace{\leftrightarrow}_{\substack{Bayes\ Theorem\ on\\ P(A\cap B|C),\ P(A|C)}} \frac{P(A\cap B\cap C)}{P(C)} = \frac{P(A\cap C)}{P(C)}P(B|C)$

$$\underbrace{\leftrightarrow}_{multiply\ by\ P(C)} P(A\cap B\cap C) = P(A\cap C)P(B|C)$$

$$\underbrace{\leftrightarrow}_{divide\ by\ P(A\cap C)>0} \frac{P(B\cap A\cap C)}{P(A\cap C)} = P(B|C) \underbrace{\leftrightarrow}_{\substack{Bayes\ Theorem\ on\\ P(B|A,C)}} P(B|A,C) = P(B|C)$$

Problem 21: Show that $E(\mathbb{1}_A) = P(A)$

Proof:

$$E(\mathbb{1}_A) = 1 * P(\mathbb{1}_A = 1) + 0 * p(\mathbb{1}_A = 0) = 1 * P(\omega \in A) + 0 * p(\omega \notin A) = P(\omega \in A) = P(A)$$

Problem 22: prove Fact 13:

1) If $X \in R^n$ and $Y \in R^m$ are orthogonal RVs, then the correlation matrix of $Z = [X^T\ Y^T]^T$ is $\begin{bmatrix} R_X & 0_{n\times m} \\ 0_{m\times n} & R_Y \end{bmatrix}$
2) If $X \in R^n$ and $Y \in R^m$ are uncorrelated RVs, then the covariance matrix of $Z = [X^T\ Y^T]^T$ is $\begin{bmatrix} \Sigma_X & 0_{n\times m} \\ 0_{m\times n} & \Sigma_Y \end{bmatrix}$

Proof:

1) $X \in R^n$ and $Y \in R^m$ are orthogonal RVs, hence from definition 37:

$$E(XY^T) = 0_{n\times m}$$

Recall from definition 28 that

$$E(XY^T) = R_{XY}, \qquad R_{YX} = R_{XY}^T$$
$$\rightarrow R_{XY} = 0_{n\times m}, \qquad R_{YX} = 0_{m\times n}$$

Finally, we saw in fact 9 that the correlation matrix of Z is $R_Z = \begin{bmatrix} R_X & \mathrm{R_{XY}} \\ \mathrm{R_{YX}} & R_Y \end{bmatrix}$

$$\rightarrow R_Z = \begin{bmatrix} R_X & \mathrm{R_{XY}} \\ \mathrm{R_{YX}} & R_Y \end{bmatrix} = \begin{bmatrix} R_X & 0_{n\times m} \\ 0_{m\times n} & R_Y \end{bmatrix}$$

2) $X \in R^n$ and $Y \in R^m$ are uncorrelated RVs, hence from definition 38: $X - \mu_X$ *and* $Y - \mu_Y$ are orthogonal, and again from definition 37 we obtain:

$$E\big((X - \mu_X)(Y - \mu_Y)^T\big) = 0_{n\times m}$$

Recall from definition 28 and problem 17 that

$$E\big((X - \mu_X)(Y - \mu_Y)^T\big) = \Sigma_{XY}, \qquad \Sigma_{YX} = \Sigma_{XY}^T$$
$$\rightarrow \Sigma_{XY} = 0_{n\times m}, \qquad \Sigma_{YX} = 0_{m\times n}$$

Finally, we saw in fact 9 that the covariance matrix of Z is $\Sigma_Z = \begin{bmatrix} \Sigma_X & \Sigma_{\mathrm{XY}} \\ \Sigma_{\mathrm{YX}} & \Sigma_Y \end{bmatrix}$

$$\rightarrow \Sigma_Z = \begin{bmatrix} \Sigma_X & \Sigma_{\mathrm{XY}} \\ \Sigma_{\mathrm{YX}} & \Sigma_Y \end{bmatrix} = \begin{bmatrix} \Sigma_X & 0_{n\times m} \\ 0_{m\times n} & \Sigma_Y \end{bmatrix}$$

Problem 23: prove Fact 14:

$$X \perp\!\!\!\perp Y \rightarrow X \perp Y$$

Proof:

X and Y are independent RVs, hence by definition $p(x,y) = p(x)p(y)\ \forall x, y$.
We need to show that X and Y are uncorrelated, meaning: $E(XY^T) = \mu_X \mu_Y^T$

For $X, Y \in R^n$ discrete RVs:

$$E(XY^T) = \sum_x \sum_y xy^T p(x,y) = \sum_x \sum_y xy^T p(x)p(y) = \sum_x xp(x) \sum_y y^T p(y) = \mu_X \mu_Y^T$$

For $X, Y \in R^n$ continuous RVs:

$$E(XY^T) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy^T p(x,y)\, dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy^T p(x)p(y)\ dxdy = \int_{-\infty}^{\infty} xp(x)dx \int_{-\infty}^{\infty} y^T p(y)dy = \mu_X \mu_Y^T$$

Problem 24: Let $X = [X_1 \;\; X_2]^T$ be a two-dimensional random vector taking values in $S = \{0,1\}^2$. The probability mass function (pmf) of X is given by:

$$p(0,0) = 0.5$$
$$p(0,1) = 0.1$$
$$p(1,0) = 0.3$$
$$p(1,1) = 0.1$$

(a) Find the marginals, $p(X1), p(X_2)$.
(b) Find the 2D mean vector, $E(X)$.
(c) Find the 2-by-2 correlation matrix, $E(XX^T)$.
(d) Find the 2-by-2 covariance matrix, $E\big((X-\mu)(X-\mu)^T\big), \;\; \mu = E(X)$.
(e) Are X_1 and X_2 independent?
(f) Are X_1 and X_2 correlated?

Proof:

(a)

$$p(X_1 = 0) = \sum_{x_2 \in \{0,1\}} p(0, x_2) = p(0,0) + p(0,1) = 0.5 + 0.1 = 0.6$$
$$p(X_1 = 1) = \sum_{x_2 \in \{0,1\}} p(1, x_2) = p(1,0) + p(1,1) = 0.3 + 0.1 = 0.4$$
$$\rightarrow p(x_1) = \begin{cases} 0.6, & x_1 = 0 \\ 0.4, & x_1 = 1 \\ 0, & otherwise \end{cases}$$

$$p(X_2 = 0) = \sum_{x_1 \in \{0,1\}} p(x_1, 0) = p(0,0) + p(1,0) = 0.5 + 0.3 = 0.8$$
$$p(X_2 = 1) = \sum_{x_1 \in \{0,1\}} p(x_1, 1) = p(0,1) + p(1,1) = 0.1 + 0.1 = 0.2$$
$$\rightarrow p(x_2) = \begin{cases} 0.8, & x_2 = 0 \\ 0.2, & x_2 = 1 \\ 0, & otherwise \end{cases}$$

(b) $E(X) = E([X_1 \;\; X_2]^T) = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} 0 * p(X_1 = 0) + 1 * p(X_1 = 1) \\ 0 * p(X_2 = 0) + 1 * p(X_2 = 1) \end{bmatrix} = \begin{bmatrix} p(X_1 = 1) \\ p(X_2 = 1) \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix}$

(c) $E(XX^T) = \sum_x xx^T p(x) = 0.5 * \begin{bmatrix} 0 \\ 0 \end{bmatrix} [0 \quad 0] + 0.1 * \begin{bmatrix} 0 \\ 1 \end{bmatrix} [0 \quad 1] + 0.3 * \begin{bmatrix} 1 \\ 0 \end{bmatrix} [1 \quad 0] + 0.1 * \begin{bmatrix} 1 \\ 1 \end{bmatrix} [1 \quad 1] =$
$0.5 * \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} + 0.1 * \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + 0.3 * \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 0.1 * \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.3 + 0.1 & 0.1 \\ 0.1 & 0.1 + 0.1 \end{bmatrix} = \begin{bmatrix} 0.4 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$

(d) $E\big((X-\mu)(X-\mu)^T\big) = E(XX^T) - \mu\mu^T = \begin{bmatrix} 0.4 & 0.1 \\ 0.1 & 0.2 \end{bmatrix} - \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} [0.4 \quad 0.2] = \begin{bmatrix} 0.4 & 0.1 \\ 0.1 & 0.2 \end{bmatrix} - \begin{bmatrix} 0.16 & 0.8 \\ 0.8 & 0.4 \end{bmatrix} =$
$\begin{bmatrix} 0.24 & -0.7 \\ -0.7 & -0.2 \end{bmatrix}$

(e) X_1 and X_2 are not independent since $p(0,0) = 0.5 \neq 0.48 = 0.3 * 0.8 = p(X_1 = 0) * p(X_2 = 0)$

(f) $E(X_1 X_2) = 0 * 0 * p(0,0) + 0 * 1 * p(0,1) + 1 * 0 * p(1,0) + 1 * 1 * p(1,1) = p(1,1) = 0.1$

$$E(X_1 X_2) - E(X_1)(X_2) = 0.1 - 0.4 * 0.2 = 0.02 \neq 0$$

Meaning, X_1 and X_2 are not uncorrelated $\rightarrow$ X_1 and X_2 are correlated.

Problem 25: Let X $= [X_1 \;\; X_2]^T$ be a two-dimensional random vector taking values in $S = \{0,1\}^2$. The probability mass function (pmf) of X is given by:

$$p(0,0) = 0.25$$
$$p(0,1) = 0.25$$
$$p(1,0) = 0.25$$
$$p(1,1) = 0.25$$

(a) Find the marginals, $p(X1), p(X_2)$.
(b) Find the 2D mean vector, $E(X)$.
(c) Find the 2-by-2 covariance matrix, $E\big((X-\mu)(X-\mu)^T\big),\;\; \mu = E(X)$.
(d) Are X_1 and X_2 independent?
(e) Are X_1 and X_2 uncorrelated?

Let $Y_1 = X_1 + X_2$ and let $Y_2 = |X_1 - X_2|$

(f) Find the pmf of $Y = [Y_1 \; Y_2]^T$.
(g) Find the marginal pmf of Y_1 and the marginal pmf of Y_2.
(h) Find the 2D mean vector of Y, $E(Y)$.
(i) Find the 2-by-2 covariance matrix of Y, $E\big((Y-\mu_Y)(Y-\mu_Y)^T\big),\;\; \mu_Y = E(Y)$.
(j) Are Y_1 and Y_2 independent?
(k) Are Y_1 and Y_2 uncorrelated

Proof:

(a)

$p(X_1 = 0) = \sum_{x_2\in\{0,1\}} p(0, x_2) = p(0,0) + p(0,1) = 0.25 + 0.25 = 0.5$
$p(X_1 = 1) = \sum_{x_2\in\{0,1\}} p(1, x_2) = p(1,0) + p(1,1) = 0.25 + 0.25 = 0.5$

$$\rightarrow p(x_1) = \begin{cases} 0.5, & x_1 = 0 \\ 0.5, & x_1 = 1 \\ 0, & otherwise \end{cases}$$

Similarly, we will obtain:

$$\rightarrow p(x_2) = \begin{cases} 0.5, & x_2 = 0 \\ 0.5, & x_2 = 1 \\ 0, & otherwise \end{cases}$$

(b) $E(X) = E([X_1 \;\; X_2]^T) = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} 0 * p(X_1 = 0) + 1 * p(X_1 = 1) \\ 0 * p(X_2 = 0) + 1 * p(X_2 = 1) \end{bmatrix} = \begin{bmatrix} p(X_1 = 1) \\ p(X_2 = 1) \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$

(c) $\Sigma_X = E(XX^T) - E(X)E(X^T) = \sum_x xx^T p(x) - (\sum_x xp(x)) * (\sum_x x^T p(x)) = 0.25 * \begin{bmatrix} 0 \\ 0 \end{bmatrix} [0 \quad 0] + 0.25 * \begin{bmatrix} 0 \\ 1 \end{bmatrix} [0 \quad 1] + 0.25 * \begin{bmatrix} 1 \\ 0 \end{bmatrix} [1 \quad 0] + 0.25 * \begin{bmatrix} 1 \\ 1 \end{bmatrix} [1 \quad 1] - \left(0.25 * \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0.25 * \begin{bmatrix} 0 \\ 1 \end{bmatrix} + 0.25 * \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.25 * \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) * (0.25 * [0 \quad 0] + 0.25 * [0 \quad 1] + 0.25 * [1 \quad 0] + 0.25 * [1 \quad 1]) = 0.25 * \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} + 0.25 * \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + 0.25 * \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 0.25 * \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - \left(\begin{bmatrix} 0 \\ 0.25 \end{bmatrix} + \begin{bmatrix} 0.25 \\ 0 \end{bmatrix} + * \begin{bmatrix} 0.25 \\ 0.25 \end{bmatrix}\right) * ([0 \quad 0.25] + [0.25 \quad 0] + [0.25 \quad 0.25]) = 0.25 * \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} * [0.5 \quad 0.5] = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix} - \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix} = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}$

(d) X_1 and X_2 are independent since

$$p(0,0) = 0.25 = 0.5 * 0.5 = p(X_1 = 0) * p(X_2 = 0)$$
$$p(0,1) = 0.25 = 0.5 * 0.5 = p(X_1 = 0) * p(X_2 = 1)$$
$$p(1,0) = 0.25 = 0.5 * 0.5 = p(X_1 = 1) * p(X_2 = 0)$$
$$p(1,1) = 0.25 = 0.5 * 0.5 = p(X_1 = 1) * p(X_2 = 1)$$

(e) $E(X_1X_2) = 0 * 0 * p(0,0) + 0 * 1 * p(0,1) + 1 * 0 * p(1,0) + 1 * 1 * p(1,1) = p(1,1) = 0.25$

$$E(X_1X_2) - E(X_1)(X_2) = 0.25 - 0.5 * 0.5 = 0$$

Meaning, X_1 and X_2 are uncorrelated.

(f) We can notice that $Y_1 \in \{0,1,2\}$ and $Y_2 \in \{0,1\}$

$$p(0,0) = p(Y_1 = 0, Y_2 = 0) = p(X_1 + X_2 = 0, |X_1 - X_2| = 0) = p(X_1 = 0, X_2 = 0) = 0.25$$
$$p(0,1) = p(Y_1 = 0, Y_2 = 1) = p(X_1 + X_2 = 0, |X_1 - X_2| = 1) = 0$$
$$p(1,0) = p(Y_1 = 1, Y_2 = 0) = p(X_1 + X_2 = 1, |X_1 - X_2| = 0) = 0$$
$$p(1,1) = p(Y_1 = 1, Y_2 = 1) = p(X_1 + X_2 = 1, |X_1 - X_2| = 1)$$
$$= p(X_1 = 1, X_2 = 0) + p(X_1 = 0, X_2 = 1) = 0.25 + 0.25 = 0.5$$
$$p(2,0) = p(Y_1 = 2, Y_2 = 0) = p(X_1 + X_2 = 2, |X_1 - X_2| = 0) = p(X_1 = 1, X_2 = 1) = 0.25$$
$$p(2,1) = p(Y_1 = 2, Y_2 = 1) = p(X_1 + X_2 = 2, |X_1 - X_2| = 1) = 0$$

(g)

$$p(Y_1 = 0) = \sum_{y_2 \in \{0,1\}} p(0, y_2) = p(0,0) + p(0,1) = 0.25 + 0 = 0.25$$
$$p(Y_1 = 1) = \sum_{y_2 \in \{0,1\}} p(1, y_2) = p(1,0) + p(1,1) = 0 + 0.5 = 0.5$$
$$p(Y_1 = 2) = \sum_{y_2 \in \{0,1\}} p(2, y_2) = p(2,0) + p(2,1) = 0.25 + 0 = 0.25$$
$$\rightarrow p(y_1) = \begin{cases} 0.25, & y_1 = 0 \\ 0.5, & y_1 = 1 \\ 0.25, & y_1 = 2 \\ 0, & otherwise \end{cases}$$

$$p(Y_2 = 0) = \sum_{y_1 \in \{0,1,2\}} p(y_1, 0) = p(0,0) + p(1,0) + p(2,0) = 0.25 + 0 + 0.25 = 0.5$$
$$p(Y_2 = 1) = \sum_{y_1 \in \{0,1,2\}} p(y_1, 1) = p(0,1) + p(1,1) + p(2,1) = 0 + 0.5 + 0 = 0.5$$
$$\rightarrow p(y_2) = \begin{cases} 0.5, & y_2 = 0 \\ 0.5, & y_2 = 1 \\ 0, & otherwise \end{cases}$$

(h) $E(Y) = E([Y_1 \ \ Y_2]^T) = \begin{bmatrix} E(Y_1) \\ E(Y_2) \end{bmatrix} = \begin{bmatrix} 0 * p(Y_1 = 0) + 1 * p(Y_1 = 1) + 2 * p(Y_1 = 2) \\ 0 * p(Y_2 = 0) + 1 * p(Y_2 = 1) \end{bmatrix} =$
$\begin{bmatrix} p(Y_1 = 1) + p(Y_1 = 2) \\ p(Y_2 = 1) \end{bmatrix} = \begin{bmatrix} 0.5 + 0.25 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.75 \\ 0.5 \end{bmatrix}$

(i) $E(YY^T) - E(Y)E(Y^T) = \sum_y yy^T p(y) - \left(\sum_y yp(y)\right) * \left(\sum_y y^T p(y)\right) = 0.25 * \begin{bmatrix} 0 \\ 0 \end{bmatrix} [0 \quad 0] + 0 *$
$\begin{bmatrix} 0 \\ 1 \end{bmatrix} [0 \quad 1] + 0 * \begin{bmatrix} 1 \\ 0 \end{bmatrix} [1 \quad 0] + 0.5 * \begin{bmatrix} 1 \\ 1 \end{bmatrix} [1 \quad 1] + 0.25 * \begin{bmatrix} 2 \\ 0 \end{bmatrix} [2 \quad 0] + 0 * \begin{bmatrix} 2 \\ 1 \end{bmatrix} [2 \quad 1] -$
$\left(0.25 * \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0 * \begin{bmatrix} 0 \\ 1 \end{bmatrix} + 0 * \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.5 * \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0.25 * \begin{bmatrix} 2 \\ 0 \end{bmatrix} + 0 * \begin{bmatrix} 2 \\ 1 \end{bmatrix}\right) (0.25 * [0 \quad 0] + 0 * [0 \quad 1] + 0 *$
$[1 \quad 0] + 0.5 * [1 \quad 1] + 0.25 * [2 \quad 0] + 0 * [2 \quad 1]) = 0.25 * \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} + 0.5 * \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + 0.25 * \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} -$
$\left(\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}\right) * ([0.5 \quad 0.5] + [0.5 \quad 0]) = \begin{bmatrix} 0.5 + 4 * 0.25 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} - \left(\begin{bmatrix} 1 \\ 0.5 \end{bmatrix} * [1 \quad 0.5]\right) = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} -$
$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 0.25 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.25 \end{bmatrix}$

(j) Y_1 and Y_2 are not independent since $p(0,0) = 0.25 \neq 0.125 = 0.25 * 0.5 = p(Y_1 = 0) * p(Y_2 = 0)$

(k) $E(Y_1 Y_2) = 0 * 0 * p(0,0) + 1 * 0 * p(1,0) + 2 * 0 * p(2,0) + 0 * 1 * p(0,1) + 1 * 1 * p(1,1) + 2 * 1 *$
$p(2,1) = p(1,1) + 2 * p(2,1) = 0.5 + 2 * 0 = 0.5$

$$E(Y_1 Y_2) - E(Y_1)(Y_2) = 0.5 - 0.75 * 0.5 = 0.5 - 0.375 = 0.125 \neq 0$$

Meaning, X_1 and X_2 are not uncorrelated $\rightarrow$ X_1 and X_2 are correlated.

Problem 26: $X \perp\!\!\!\perp Y|Z \leftrightarrow p(x,y,z) = p(x|z) \leftrightarrow p(y|x,z) = p(y|z)$

Proof: We'll prove:

1) $X \perp\!\!\!\perp Y|Z \leftrightarrow p(x|y,z) = p(x|z)$
2) $X \perp\!\!\!\perp Y|Z \leftrightarrow p(y|x,z) = p(y|z)$

Let us notice that:

$$(*) \quad p(x,y|z) \underbrace{=}_{\substack{Bayes\\Theorem}} \frac{p(x,y,z)}{p(z)} \underbrace{=}_{Definition\ 40} p(x|z)p(y|z) \underbrace{\leftrightarrow}_{\substack{multiply\\by\ p(z)}} p(x,y,z) = p(x|z)p(y|z)\,p(z)$$

1) $$X \perp\!\!\!\perp Y|Z \leftrightarrow p(x|y,z) \underbrace{=}_{\substack{Bayes\\Theorem}} \frac{p(x,y,z)}{p(y,z)} \underbrace{=}_{from\ (*)} \frac{p(x|z)p(y|z)p(z)}{p(y,z)} \underbrace{=}_{\substack{Bayes\\Theorem}} \frac{p(x|z)\frac{p(y,z)}{p(z)}p(z)}{p(y,z)} = p(x|z)$$
2) $$X \perp\!\!\!\perp Y|Z \leftrightarrow p(y|x,z) \underbrace{=}_{\substack{Bayes\\Theorem}} \frac{p(x,y,z)}{p(x,z)} \underbrace{=}_{from\ (*)} \frac{p(x|z)p(y|z)p(z)}{p(x,z)} \underbrace{=}_{\substack{Bayes\\Theorem}} \frac{\frac{p(x,z)}{p(z)}p(y|z)p(z)}{p(x,z)} = p(y|z)$$

Problem 27: Let B be some bounded hyper-rectangle in R^3.

1) Let $p(x,y,z) \propto \begin{cases} \exp(x+xz+yz) & if\ (x,y,z) \in B \\ 0 & otherwise \end{cases}$

   Show that $X \perp\!\!\!\perp Y|Z$

2) Let $p(x,y,z) \propto \begin{cases} \exp(xyz) & if\ (x,y,z) \in B \\ 0 & otherwise \end{cases}$

   Show that $X \not\perp\!\!\!\perp Y|Z$

Proof:

1) For $f(x,z) = \begin{cases} \exp(x+xz) & if\ (x,y,z) \in B\ for\ some\ y \\ 0 & otherwise \end{cases}$

   and $g(y,z) = \begin{cases} \exp(yz) & if\ (x,y,z) \in B\ \ for\ some\ x \\ 0 & otherwise \end{cases}$

   we get that:

   $$p(x,y,z) \propto \begin{cases} \exp(x+xz+yz) & if\ (x,y,z) \in B \\ 0 & otherwise \end{cases} = \begin{cases} f(x,z) * g(y,z) & if\ (x,y,z) \in B \\ 0 & otherwise \end{cases}$$

   and now from Fact 18 we get that: $X \perp\!\!\!\perp Y|Z$.

2) For bounded hyper-rectangle $B = \{(i,j,k): 0 \le i,j,k \le 1\}$ we can see that for $x = 1, y = 0.5, z = 1$:

   $$p(x,y|z) = \frac{p(x,y,z)}{p(z)} = \frac{p(x,y,z)}{\iint p(x,y,z)\,dxdy} \propto \frac{\exp(1*0.5*1)}{\iint \exp(xy*1)\,dxdy} = \frac{\exp(0.5)}{\iint \exp(xy)\,dxdy} \approx \frac{1.6487}{\mathbf{1.3179}}$$
   $$\approx 1.251$$
   $$p(x|z)p(y|z) = \frac{p(x,z)}{p(z)} * \frac{p(y,z)}{p(z)} = \frac{\int p(x,y,z)dy}{\iint p(x,y,z)\,dxdy} * \frac{\int p(x,y,z)\,dx}{\iint p(x,y,z)\,dxdy}$$
   $$\propto \frac{\int \exp(1*y*1)\,dy}{\iint \exp(xy*1)\,dxdy} * \frac{\int \exp(x*0.5*1)\,dx}{\iint \exp(xy*1)\,dxdy} = \frac{\int \exp(y)\,dy}{\iint \exp(xy)\,dxdy} * \frac{\int \exp(0.5x)\,dx}{\iint \exp(xy)\,dxdy}$$
   $$= \frac{\int \exp(y)\,dy}{\iint \exp(xy)\,dxdy} * \frac{\int \exp(0.5x)\,dx}{\iint \exp(xy)\,dxdy} \approx \frac{(\mathrm{e}-1)}{\mathbf{1.3179}} * \frac{2\sqrt{e}-2}{\mathbf{1.3179}} \approx 1.283$$

   And: $1.251 \neq 1.283$

   Therefore: $X \not\perp\!\!\!\perp Y|Z$.

Problem 28: Show that $\mathcal{P}(\Omega)$ is σ-field

<u>Proof</u>:
We will show that:

1) $\Omega \in \mathcal{P}(\Omega)$
2) $\mathcal{P}(\Omega)$ is closed under complementation
3) $\mathcal{P}(\Omega)$ is closed under countable unions

    1) From definition of power-set: $\Omega \in \mathcal{P}(\Omega)$
    2) Mark $\{\sigma_1, \sigma_2, \ldots\} \in \Omega$. For $A = \{\sigma_i : for\ some\ i's\} \in \Omega$, $A^c = \Omega \backslash A = \{\sigma_j : j \neq i\}$.
       Now, from power-set definition we get that: $\{\sigma_j : j \neq i\} \in \mathcal{P}(\Omega)$.
    3) Mark $\{\sigma_1, \sigma_2, \ldots\} \in \Omega$. For $A_i = \left\{\sigma_j^{(i)} : for\ some\ j's\right\} \in \Omega$
       $\bigcup_{n=1}^{\infty} A_n = \bigcup\left\{\sigma_j : \sigma_j^{(i)} \in A_i\right\}$ Now, from power-set definition we get that: $\bigcup\left\{\sigma_j : \sigma_j^{(i)} \in A_i\right\} \in \mathcal{P}(\Omega)$.

Problem 29: Let $\mathcal{F}$ be a σ-field.

1) Show that $\emptyset \in \mathcal{F}$.
2) Using De Morgan's laws, show that $\mathcal{F}$ is closed under countable intersections. Namely, if $A_1, A_2, \ldots$ are in $\mathcal{F}$ then $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$

<u>Proof</u>:

1) Because $\Omega \in \mathcal{F}$, we get from property (2) of σ-field that: $\Omega^c = \emptyset \in \mathcal{F}$
2) De Morgan's laws state that $(A \cup B)^c = A^c \cap B^c$.Therefore: $\bigcup_i A_i = \bigcap_i A_i^c$
   For all $A_i \in \mathcal{F}$, from property (2) of $\sigma -$ field: $A_i^c \in \mathcal{F}$
   from property (3) of $\sigma -$ field: $\bigcup_i A_i^c \in \mathcal{F}$, and therefore: $\bigcup_i A_i^c = \bigcap_i (A_i^c)^c = \bigcap_i A_i \in \mathcal{F}$.