Homework 1

Problem 1. Consider the tossing a coin experiment. Suppose that $\Pr(H) = \Pr(T) = 1/2$. Define random variables X and Y as follows: $X(H) = 1$, $X(T) = 2$, $Y(H) = 2$, $Y(T) = 1$. Find CDFs of X and Y. Are X and Y equal in distribution (do they have the same CDFs.)? What is $\Pr(X = Y)$?

Solution. Two random variables, X and Y, are defined as follows:

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = H; \\ 2 & \text{if } \omega = T. \end{cases}$$

$$Y(\omega) = \begin{cases} 1 & \text{if } \omega = T; \\ 2 & \text{if } \omega = H. \end{cases}$$

Therefore, the CDF of X is:

$$F_X(x) = \Pr(X \leq x) = \begin{cases} 0 & \text{if } x < 1; \\ 1/2 & \text{if } x \in [1, 2); \\ 1 & \text{if } x \geq 2. \end{cases}$$

Similarly, the CDF of Y is

$$F_Y(y) = \Pr(Y \leq y) = \begin{cases} 0 & \text{if } y < 1; \\ 1/2 & \text{if } y \in [1, 2); \\ 1 & \text{if } y \geq 2. \end{cases}$$

Thus, X and Y have the same distribution. However, $\Pr(\omega \in \Omega : X(\omega) = Y(\omega)) = \Pr(\emptyset) = 0$.

Problem 2. Suppose that the average distance between a random variable X and a constant c is measured by the function $E(X - c)^2$. (Note that $E(X - c)^2$ can be viewed as a function of c: $Q(c) = E(X - c)^2$.)

1. Show that $E(X - c)^2 = E(X - EX)^2 + (EX - c)^2$.
2. What value of c does minimize $E(X - c)^2$?

Solution. For part 1:

$$\begin{aligned} E(X - c)^2 &= E(X - EX + EX - c)^2 \\ &= E(X - EX)^2 + (EX - c)^2 + 2(EX - c)E(X - EX) \\ &= E(X - EX)^2 + (EX - c)^2, \end{aligned}$$

where the last equality holds because $E(X - EX) = 0$. For part 2, Since $E(X - EX)^2$ does not depend on c, $\min_c(E(X - EX)^2 + (EX - c)^2)$ is equivalent to $\min_c(EX - c)^2$. Since $(EX - c)^2 \geq 0$ for all $c \in \mathbb{R}$, and $(EX - c)^2 = 0$ when $c = EX$, it follows that $c = EX$ minimizes the whole expression.

Problem 3. Let X and Y be two continuously distributed random variables with the joint PDF given by

$$f_{X,Y}(x, y) = 1(0 < y < x < \sqrt{2}),$$

where $1(A)$ is the so-called *indicator function*:

$$1(A) = \begin{cases} 1, & \text{if condition } A \text{ is true}, \\ 0, & \text{if condition } A \text{ is false}. \end{cases}$$

Thus,

$$1(0 < y < x < \sqrt{2}) = \begin{cases} 1, & \text{if } 0 < y < x < \sqrt{2}, \\ 0, & \text{otherwise}. \end{cases}$$

1. Show that $f_{X,Y}$ is a PDF. Hint: Start by plotting the support of the distribution (the region where the PDF is non-zero).
2. Are X and Y statistically independent?
3. Find the marginal PDF of X.
4. Find the marginal PDF of Y.
5. Find the conditional PDF of Y given X.
6. Find $E(Y|X)$.

Solution. (1) First, $f_{X,Y}(x,y) \geq 0$ for all x's and y's. Second, $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = \int_0^{\sqrt{2}}\int_0^x dydx = \int_0^{\sqrt{2}} xdx = 1$. Hence, it is a PDF.

(2) Since $f_{X,Y}(x,y)$ cannot be written as $f_X(x)f_Y(y)$, X and Y are not independent.

(3) For $x \in (0,\sqrt{2})$, $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_0^x dy = x$. Hence, $f_X(x) = x \times 1(0 < x < \sqrt{2})$.

(4) For $y \in (0,\sqrt{2})$, $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_y^{\sqrt{2}} dx = (\sqrt{2} - y)$. Hence, $f_Y(y) = (\sqrt{2} - y) \times 1(0 < y < \sqrt{2})$.

(5) For $x \in (0,\sqrt{2})$ and $y \in (0,x)$, $f_{Y|X}(y|x) = f_{X,Y}(x,y)/f_X(x) = 1/x$. Hence, for $x \in (0,\sqrt{2})$, $f_{Y|X}(y|x) = x^{-1} \times 1(0 < y < x)$.

(6) Since the conditional distribution of Y given X is uniform on the interval $(0,X)$, $E(Y|X) = X/2$. This can be verified as follows: $\int_{-\infty}^{\infty} y f_{Y|X}(y|X)dy = X^{-1}\int_0^X ydy = X/2$.

Problem 4. Show that $E(X|Y) = E(X)$ implies that $Cov(X,Y) = 0$. Hint: apply the law of iterated expectation to show that $E(XY) = E(X)E(Y)$.

Solution. By law of iterated expectation,

$$E(XY) = E(E(XY|Y)) = E(Y \cdot E(X|Y)) = E(Y \cdot E(X)) = E(Y)E(X).$$

Problem 5. We say that X is χ^2 distributed with n degrees of freedom (denoted as $X \sim \chi^2_n$) if X can be written as $X = Z_1^2 + \ldots + Z_n^2$, where $Z_1, \ldots, Z_n$ are independent standard normal random variables. Show that $EX = n$ and $Var(X) = 2n$. Hint: Use the moment generating function to show that if $Z \sim N(0,1)$, then $EZ^4 = 3$.

Solution. First prove that if $Z \sim N(0,1)$, then $EZ^4 = 3$
we know that for $Z \sim N(0,1)$ $M(t) = e^{t^2/2}$

$$\begin{aligned} M^{(1)}(t) &= te^{t^2/2} \\ M^{(2)}(t) &= e^{t^2/2} + t^2e^{t^2/2} = M(t) + tM^{(1)}(t) \\ M^{(3)}(t) &= 2M^{(1)}(t) + tM^{(2)}(t) \\ M^{(4)}(t) &= 3M^{(2)}(t) + tM^{(3)}(t) \end{aligned}$$

Use the property of moment generating function: $M^{(s)}(0) = EZ^s \Longrightarrow EZ^4 = M^{(4)}(0) = 3e^{0^2/2} = 3$.

Then, calculate $E(X)$ and $Var(X)$. $X \sim \chi^2_n \Longrightarrow X = \sum_{i=1}^n Z_i^2$ where,$Z_i \sim N(0,1)$

$$E(X) = E(\sum_{i=1}^n Z_i^2) = \sum_{i=1}^n E(Z_i^2) = \sum_{i=1}^n Var(Z_i) = \sum_{i=1}^n 1 = n$$

$$
\begin{aligned}
E(X^2) &= E(\sum_{i=1}^{n} Z_i^2)^2 = E(\sum_{i=1}^{n}\sum_{j=1}^{n} Z_i^2 Z_j^2) \text{ there are } n^2\text{terms} \\
&= \sum_{i=1}^{n} E(Z_i^4) + \sum_{i\neq j} E(Z_i^2 Z_j^2) \\
& \quad \text{there are } n\text{terms in the first part and } n^2 - n\text{terms in the second part} \\
&= \sum_{i=1}^{n} E(Z_i^4) + \sum_{i\neq j} E(Z_i^2)E(Z_j^2) \\
& \text{because } Z_i\text{and } Z_j\text{are indepentent} \\
&= \sum_{i=1}^{n} E(Z_i^4) + \sum_{i\neq j} Var(Z_i)Var(Z_j) \\
&= \sum_{i=1}^{n} 3 + \sum_{i\neq j} 1 \\
&= 3n + n^2 - n
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow Var(X) &= E(X^2) - (EX)^2 \\
&= 3n + n^2 - n - n^2 \\
&= 2n.
\end{aligned}
$$

Problem 6. Let $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ be random vectors. Show that:

1. $Var\left(\boldsymbol{X}\right) = E\left(\boldsymbol{X}\boldsymbol{X}'\right) - E\left(\boldsymbol{X}\right)E\left(\boldsymbol{X}'\right).$
2. $Cov\left(\boldsymbol{X},\boldsymbol{Y}\right) = \left(Cov\left(\boldsymbol{Y},\boldsymbol{X}\right)\right)'.$
3. $Var\left(\boldsymbol{X}+\boldsymbol{Y}\right) = Var\left(\boldsymbol{X}\right) + Var\left(\boldsymbol{Y}\right) + Cov\left(\boldsymbol{X},\boldsymbol{Y}\right) + Cov\left(\boldsymbol{Y},\boldsymbol{X}\right).$
4. $Var(\boldsymbol{\alpha}+\boldsymbol{\Gamma}\boldsymbol{X}) = \boldsymbol{\Gamma}Var(\boldsymbol{X})\boldsymbol{\Gamma}'$, where $\boldsymbol{X}$ is a random n-vector, $\boldsymbol{\alpha}$ is a non-random k-vector, and $\boldsymbol{\Gamma}$ is a non-random $k\times n$ matrix.
5. $Cov(\mathbf{A}\boldsymbol{X}+\mathbf{B}\boldsymbol{Y},\mathbf{C}\boldsymbol{Z}) = \mathbf{A}(Cov(\boldsymbol{X},\boldsymbol{Z}))\mathbf{C}' + \mathbf{B}(Cov(\boldsymbol{Y},\boldsymbol{Z}))\mathbf{C}'$, where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are non-random matrices.

Solution. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two random vectors.

$$
\begin{aligned}
1)\ Var(\boldsymbol{X}) \quad &= E(\boldsymbol{X}-E(\boldsymbol{X}))(\boldsymbol{X}-E(\boldsymbol{X}))' \\
&= E(\boldsymbol{X}\boldsymbol{X}' - \boldsymbol{X}\cdot E(\boldsymbol{X}') - E(\boldsymbol{X})\boldsymbol{X}' + E(\boldsymbol{X})E(\boldsymbol{X}')) \\
&= E(\boldsymbol{X}\boldsymbol{X}') - E(\boldsymbol{X})E(\boldsymbol{X}') - E(\boldsymbol{X})E(\boldsymbol{X}') + E(\boldsymbol{X})E(\boldsymbol{X}') \\
&= E(\boldsymbol{X}\boldsymbol{X}') - E(\boldsymbol{X})E(\boldsymbol{X}').
\end{aligned}
$$

$$
\begin{aligned}
2)\ Cov(\boldsymbol{X},\boldsymbol{Y}) \quad &= E(\boldsymbol{X}\boldsymbol{Y}') - E(\boldsymbol{X})E(\boldsymbol{Y}') \\
& \quad \text{using conclusion in part (1)} \\
&= [E(\boldsymbol{Y}\boldsymbol{X}') - E(\boldsymbol{Y})E(\boldsymbol{X}')]' \\
&= Cov(\boldsymbol{Y},\boldsymbol{X})'.
\end{aligned}
$$

$$
3)\ Var(\boldsymbol{X}+\boldsymbol{Y}) \quad = E(\boldsymbol{X}+\boldsymbol{Y})(\boldsymbol{X}+\boldsymbol{Y})' - E(\boldsymbol{X}+\boldsymbol{Y})E(\boldsymbol{X}+\boldsymbol{Y})'
$$

$$
\begin{aligned}
&\text{using conclusion in part (1)}\\
=&E(\boldsymbol{X}\boldsymbol{X}' + \boldsymbol{X}\boldsymbol{Y}' + \boldsymbol{Y}\boldsymbol{X}' + \boldsymbol{Y}\boldsymbol{Y}') - E(\boldsymbol{X})E(\boldsymbol{X}')\\
&- E(\boldsymbol{X})E(\boldsymbol{Y}') - E(\boldsymbol{Y})E(\boldsymbol{X}') - E(\boldsymbol{Y})E(Y')\\
=&[E(\boldsymbol{X}\boldsymbol{X}') - E(\boldsymbol{X})E(\boldsymbol{X}')] + [E(\boldsymbol{Y}\boldsymbol{Y}') - E(\boldsymbol{Y})E(\boldsymbol{Y}')]\\
&+ [E(\boldsymbol{X}\boldsymbol{Y}') - E(\boldsymbol{X})E(\boldsymbol{Y}')] + [E(\boldsymbol{Y}\boldsymbol{X}') - E(\boldsymbol{Y})E(\boldsymbol{X}')]\\
=&Var(\boldsymbol{X}) + Var(\boldsymbol{Y}) + Cov(\boldsymbol{X},\boldsymbol{Y}) + Cov(\boldsymbol{Y},\boldsymbol{X}).
\end{aligned}
$$

$$
\begin{aligned}
4)Var(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{X}) \quad =&E[\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{X} - E(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{X})][\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{X} - E(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{X})]'\\
=&E[\boldsymbol{\Gamma}\boldsymbol{X} - E(\boldsymbol{\Gamma}\boldsymbol{X})][\boldsymbol{\Gamma}\boldsymbol{X} - E(\boldsymbol{\Gamma}\boldsymbol{X})]'\\
=&E\boldsymbol{\Gamma}(\boldsymbol{X} - E\boldsymbol{X})[\boldsymbol{\Gamma}(\boldsymbol{X} - E\boldsymbol{X})]'\\
=&\boldsymbol{\Gamma}E[(\boldsymbol{X} - E\boldsymbol{X})(\boldsymbol{X} - E\boldsymbol{X})']\boldsymbol{\Gamma}'\\
=&\boldsymbol{\Gamma}Var(\boldsymbol{X})\boldsymbol{\Gamma}'.
\end{aligned}
$$

$$
\begin{aligned}
5)Cov(\mathbf{A}\boldsymbol{X} + \mathbf{B}\boldsymbol{Y}, \mathbf{C}\boldsymbol{Z}) &= E[(\mathbf{A}\boldsymbol{X} + \mathbf{B}\boldsymbol{Y})\boldsymbol{Z}'\mathbf{C}'] - E(\mathbf{A}\boldsymbol{X} + \mathbf{B}\boldsymbol{Y})E(\boldsymbol{Z}'\mathbf{C}')\\
&= E(\mathbf{A}\boldsymbol{X}\boldsymbol{Z}'\mathbf{C}') - E(\mathbf{A}\boldsymbol{X})E(\boldsymbol{Z}'\mathbf{C}') + E(\mathbf{B}\boldsymbol{Y}\boldsymbol{Z}'\mathbf{C}') - E(\mathbf{B}\boldsymbol{Y})E(\boldsymbol{Z}'\mathbf{C}')\\
&= \mathbf{A}[E(\boldsymbol{X}\boldsymbol{Z}') - E(\boldsymbol{X})E(\boldsymbol{Z}')]\mathbf{C}' + \mathbf{B}[E(\boldsymbol{Y}\boldsymbol{Z}') - E(\boldsymbol{Y})E(\boldsymbol{Z}')]\mathbf{C}'\\
&\quad \text{because } \mathbf{A}, \mathbf{B} \text{ and } \mathbf{C} \text{ are non-random}\\
&= \mathbf{A}(Cov(\boldsymbol{X},\boldsymbol{Z}))\mathbf{C}' + \mathbf{B}(Cov(\boldsymbol{Y},\boldsymbol{Z}))\mathbf{C}'.
\end{aligned}
$$

Problem 7. The Mean Trimmed Squared Error (MTSE) is defined by

$$T(\boldsymbol{\theta}) = \mathrm{E}\left(\left(Y - \boldsymbol{X}'\boldsymbol{\theta}\right)^2 \tau(\boldsymbol{X})\right),$$

where $\tau(\boldsymbol{X})$ is a known, scalar-valued, non-negative, bounded, function.

1. Give an explicit formula for the value of $\boldsymbol{\theta}$ which minimizes $T(\boldsymbol{\theta})$.
2. Define $e = Y - \boldsymbol{X}'\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is the minimizer defined above. Show: $\mathrm{E}(\boldsymbol{X}\tau(\boldsymbol{X})e) = 0$.
3. Under what condition (other than $\tau(\boldsymbol{X}) = 1$) will this minimizer equal the Best Linear Predictor?

Solution. Part (1). By expanding the square

$$
\begin{aligned}
T(\boldsymbol{\theta}) &= \mathbb{E}\left(\left(Y - \boldsymbol{X}'\boldsymbol{\theta}\right)^2 \tau(\boldsymbol{X})\right)\\
&= \mathbb{E}\left(Y^2\tau(\boldsymbol{X})\right) - 2\mathbb{E}\left(Y\boldsymbol{X}'\tau(\boldsymbol{X})\right)\boldsymbol{\theta} + \boldsymbol{\theta}'\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau(\boldsymbol{X})\right)\boldsymbol{\theta}.
\end{aligned}
$$

Differentiate:

$$\frac{\partial}{\partial\boldsymbol{\theta}}T(\boldsymbol{\theta}) = -2\mathbb{E}(\boldsymbol{X}Y\tau(\boldsymbol{X})) + 2\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau(\boldsymbol{X})\right)\boldsymbol{\theta}.$$

Setting it equal to zero and solving for $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = \left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau(\boldsymbol{X})\right)\right)^{-1}\mathbb{E}(\boldsymbol{X}Y\tau(\boldsymbol{X})).$$

Part (2). Since $e = Y - \boldsymbol{X}'\boldsymbol{\theta}$,

$$
\begin{aligned}
\mathbb{E}(\boldsymbol{X}e\tau(\boldsymbol{X})) &= \mathbb{E}(\boldsymbol{X}Y\tau(\boldsymbol{X})) - \mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau(\boldsymbol{X})\right)\boldsymbol{\theta}\\
&= \mathbb{E}(\boldsymbol{X}Y\tau(\boldsymbol{X})) - \mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau(\boldsymbol{X})\right)\left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau(\boldsymbol{X})\right)\right)^{-1}\mathbb{E}(\boldsymbol{X}Y\tau(\boldsymbol{X}))\\
&= 0.
\end{aligned}
$$

Part (3). If the conditional mean is linear: $\mathbb{E}(Y|\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{\beta}$ then by the law of iterated expectation,

$$\begin{aligned}
\boldsymbol{\theta} &= \left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau(\boldsymbol{X})\right)\right)^{-1}\mathbb{E}\left(\boldsymbol{X}Y\tau(\boldsymbol{X})\right) \\
&= \left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau(\boldsymbol{X})\right)\right)^{-1}\mathbb{E}\mathbb{E}\left(\boldsymbol{X}Y\tau(\boldsymbol{X})|\boldsymbol{X}\right) \\
&= \left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau(\boldsymbol{X})\right)\right)^{-1}\mathbb{E}\boldsymbol{X}\tau(\boldsymbol{X})\mathbb{E}(Y|\boldsymbol{X}) \\
&= \left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\tau(\boldsymbol{X})\right)\right)^{-1}\mathbb{E}\left(\boldsymbol{X}\tau(\boldsymbol{X})\boldsymbol{X}'\right)\boldsymbol{\beta} \\
&= \boldsymbol{\beta}.
\end{aligned}$$

Problem 8. Suppose that

$$\begin{aligned}
Y &= \boldsymbol{X}'\boldsymbol{\beta} + e \\
\mathrm{E}(e|\boldsymbol{X}) &= 0 \\
\mathrm{E}\left(e^2|\boldsymbol{X}\right) &= \sigma^2(\boldsymbol{X}).
\end{aligned}$$

Consider two approximations to the conditional variance $\sigma^2(\boldsymbol{X})$:

$$\boldsymbol{\gamma}_1 \text{ minimizes } \mathrm{E}\left(\sigma^2(\boldsymbol{X}) - \boldsymbol{X}'\boldsymbol{\gamma}\right)^2$$

and

$$\boldsymbol{\gamma}_2 \text{ minimizes } \mathrm{E}\left(e^2 - \boldsymbol{X}'\boldsymbol{\gamma}\right)^2.$$

Show: $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2$.

Solution. By law of iterated expectation,

$$\begin{aligned}
\boldsymbol{\gamma}_2 &= \left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\right)\right)^{-1}\mathbb{E}\left(\boldsymbol{X}e^2\right) \\
&= \left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\right)\right)^{-1}\mathbb{E}\left(\boldsymbol{X}\mathbb{E}\left(e^2|\boldsymbol{X}\right)\right) \\
&= \left(\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\right)\right)^{-1}\mathbb{E}\left(\boldsymbol{X}\sigma^2(\boldsymbol{X})\right) \\
&= \boldsymbol{\gamma}_1.
\end{aligned}$$

Problem 9. Suppose that you had a new battery for your camera, and the life of the battery is a random variable X, with PDF

$$f_X(x) = k \times \exp\left(-\frac{x}{\beta}\right),$$

where $x > 0$ and β is a parameter. Assume now that t and s are non-negative real numbers.

(a). Use the properties of a PDF to determine the value of k.

(b). Find an expression for $\Pr(X \geq t)$.

(c). Find an expression for the conditional probability: $\Pr(X \geq t + s|X \geq s)$. Hint: Use $\Pr(A|B) = \frac{\Pr(A\cap B)}{\Pr(B)}$.

(d). Suppose that your battery has already lasted for s weeks without dying. Based on your above answers, are you more concerned that the battery is about to die than you were when you first put it in the camera?

Solution. (a) We know that the PDF must integrate to 1:

$$\begin{aligned}
\int_0^\infty f_X(x)\,\mathrm{d}x &= \int_0^\infty k \times \exp\left(-\frac{x}{\beta}\right)\mathrm{d}x \\
1 &= k\int_0^\infty \exp\left(-\frac{x}{\beta}\right)\mathrm{d}x \\
\frac{1}{k} &= -\beta \cdot \exp\left(-\frac{x}{\beta}\right)\Big|_0^\infty
\end{aligned}$$

$$k = \frac{1}{\beta}.$$

(b) This expression is straightforward now that we have the integration constant, k:

$$P\left(X \geq t\right) = \int_t^\infty \frac{1}{\beta} \exp\left(-{}^x\!/\!_\beta\right) \mathrm{d}x$$

which can be simplified to

$$P\left(X \geq t\right) = \exp\left(-{}^t\!/\!_\beta\right).$$

(c)

$$P\left(X \geq t + s | X \geq s\right) = \frac{\exp\left(-{}^{(t+s)}\!/\!_\beta\right)}{\exp\left(-{}^s\!/\!_\beta\right)} = \exp\left(-{}^t\!/\!_\beta\right).$$

(d) If my batteries have lasted s weeks without dying, based on my answer to part 3, I should be just as worried as I was before, since survival of the battery tells me nothing new about its likelihood of dying. The exponential distribution (which this is) has this very special property, that no matter how long something has lasted, its rate/probability of failure is constant at any given time.

Problem 10. If $\mathbf{A}$ is a symmetric positive definite $k \times k$ matrix, then $\mathbf{I} - \mathbf{A}$ is positive definite if and only if $\mathbf{A}^{-1} - \mathbf{I}$ is positive definite, where $\mathbf{I}$ is the $k \times k$ identity matrix. Prove this result by considering the quadratic form $\boldsymbol{x}'\left(\mathbf{I} - \mathbf{A}\right)\boldsymbol{x}$ and expressing $\boldsymbol{x}$ as $\mathbf{R}^{-1}\boldsymbol{z}$, where $\mathbf{R}$ is a symmetric matrix such that $\mathbf{A} = \mathbf{R}^2$. Then, extend this result to show that if $\mathbf{A}$ and $\mathbf{B}$ are symmetric positive definite matrices of the same dimensions, then $\mathbf{A} - \mathbf{B}$ is positive definite if and only if $\mathbf{B}^{-1} - \mathbf{A}^{-1}$ is positive definite. **Hint:** $\mathbf{A}$ is symmetric and positive definite, and therefore it can be written as $\mathbf{A} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}'$, where $\boldsymbol{\Lambda}$ is the diagonal matrix composed of the positive eigenvalues, $\mathbf{C}$ is the matrix of eigenvectors, and $\mathbf{C}'\mathbf{C} = \mathbf{I}$. Now, one can write $\mathbf{A} = \mathbf{R}\mathbf{R}$, where $\mathbf{R} = \mathbf{C}\boldsymbol{\Lambda}^{1/2}\mathbf{C}'$. Furthermore, $\mathbf{A}^{-1} = \mathbf{R}^{-1}\mathbf{R}^{-1}$.

Solution. (a) $\mathbf{A}$ is a symmetric positive definite $n \times n$ matrix.

Claim: $\mathbf{I} - \mathbf{A}$ is positive definite $\Leftrightarrow$ $\mathbf{A}^{-1} - \mathbf{I}$ is positive definite.

First, we need to show that $\mathbf{I} - \mathbf{A}$ is positive definite $\Rightarrow$ $\mathbf{A}^{-1} - \mathbf{I}$ is positive definite. Since $\mathbf{A}$ is symmetric and positive definite, $\mathbf{A}$ can be written as $\mathbf{A} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}'$ where $\boldsymbol{\Lambda}$ is a diagonal matrix of positive eigenvalues, $\mathbf{C}\mathbf{C}' = \mathbf{C}'\mathbf{C} = \mathbf{I}$ and $\mathbf{C}$ is the matrix of eigenvectors. One can write $\mathbf{A} = \mathbf{R}\mathbf{R}$, where $\mathbf{R} = \mathbf{C}\boldsymbol{\Lambda}^{1/2}\mathbf{C}'$. Also, $\mathbf{A}^{1/2} = \mathbf{C}\boldsymbol{\Lambda}^{1/2}\mathbf{C}'$, $\mathbf{A}^{-1/2} = \mathbf{C}\boldsymbol{\Lambda}^{-1/2}\mathbf{C}'$, $\mathbf{A}^{-1} = \mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}'$ and $\mathbf{C}' = \mathbf{C}^{-1}$.

$\mathbf{I} - \mathbf{A}$ positive definite means that for any $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x} \neq \mathbf{0}$, $\boldsymbol{x}'(\mathbf{I} - \mathbf{A})\boldsymbol{x} > 0$. Let $\boldsymbol{x} = \mathbf{R}^{-1}\boldsymbol{z} = \mathbf{A}^{-1/2}\boldsymbol{z}$ where $\boldsymbol{z} \in \mathbb{R}^n,\ \boldsymbol{z} \neq \mathbf{0}$. Notice that $\mathbf{A}^{-1/2} = \left(\mathbf{A}^{-1/2}\right)'$ since $\boldsymbol{\Lambda}^{-1/2} = \left(\boldsymbol{\Lambda}^{-1/2}\right)'$ ($\boldsymbol{\Lambda}$ is a diagonal matrix). Then, $\boldsymbol{x}'(\mathbf{I} - \mathbf{A})\boldsymbol{x} = \boldsymbol{z}'\left(\mathbf{A}^{-1/2}\right)'(\mathbf{I} - \mathbf{A})\mathbf{A}^{-1/2}\boldsymbol{z} = \boldsymbol{z}'\mathbf{A}^{-1/2}(\mathbf{I} - \mathbf{A})\mathbf{A}^{-1/2}\boldsymbol{z} = \boldsymbol{z}'(\mathbf{A}^{-1} - \mathbf{I})\boldsymbol{z} > 0$. Since $\boldsymbol{z}$ is an arbitrary vector, where $\boldsymbol{z} \in \mathbb{R}^n,\ \boldsymbol{z} \neq \mathbf{0}$, the result follows.

Second, we need to show that $\mathbf{A}^{-1} - \mathbf{I}$ positive definite $\Rightarrow \mathbf{I} - \mathbf{A}$ is positive definite. $\mathbf{A}^{-1} - \mathbf{I}$ positive definite means that for any $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{y} \neq \mathbf{0}$, $\boldsymbol{y}'(\mathbf{A}^{-1} - \mathbf{I})\boldsymbol{y} > 0$. Let $\boldsymbol{y} = \mathbf{R}\boldsymbol{w} = \mathbf{A}^{1/2}\boldsymbol{w}$ where $\boldsymbol{w} \in \mathbb{R}^n,\ \boldsymbol{w} \neq \mathbf{0}$. Notice that $\mathbf{A}^{1/2} = \left(\mathbf{A}^{1/2}\right)'$ since $\boldsymbol{\Lambda}^{1/2} = \left(\boldsymbol{\Lambda}^{1/2}\right)'$ ($\boldsymbol{\Lambda}$ is a diagonal matrix). Then, $\boldsymbol{y}'(\mathbf{I} - \mathbf{A})\boldsymbol{y} = \boldsymbol{w}'\left(\mathbf{A}^{1/2}\right)'(\mathbf{A}^{-1} - \mathbf{I})\mathbf{A}^{1/2}\boldsymbol{w} = \boldsymbol{w}'\mathbf{A}^{1/2}(\mathbf{A}^{-1} - \mathbf{I})\mathbf{A}^{1/2}\boldsymbol{w} = \boldsymbol{w}'(\mathbf{I} - \mathbf{A})\boldsymbol{w} > 0$. Since $\boldsymbol{w}$ is an arbitrary vector, where $\boldsymbol{w} \in \mathbb{R}^n,\ \boldsymbol{w} \neq \mathbf{0}$, the result follows.

(b) $\mathbf{A}$ and $\mathbf{B}$ are symmetric positive definite matrices of the same dimension.

Claim: $\mathbf{A} - \mathbf{B}$ is positive definite $\Leftrightarrow$ $\mathbf{B}^{-1} - \mathbf{A}^{-1}$ is positive definite.

First, we need to show that $\mathbf{A} - \mathbf{B}$ is positive definite $\Rightarrow$ $\mathbf{B}^{-1} - \mathbf{A}^{-1}$ is positive definite. $\mathbf{A} - \mathbf{B}$ positive definite means that for any $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x} \neq \mathbf{0}$, $\boldsymbol{x}'(\mathbf{A} - \mathbf{B})\boldsymbol{x} > 0$. Let $\boldsymbol{x} = \mathbf{R}^{-1}\boldsymbol{z} = \mathbf{A}^{-1/2}\boldsymbol{z}$ where $\boldsymbol{z} \in \mathbb{R}^n,\ \boldsymbol{z} \neq \mathbf{0}$. Then, $\boldsymbol{x}'(\mathbf{A} - \mathbf{B})\boldsymbol{x} = \boldsymbol{z}'\left(\mathbf{A}^{-1/2}\right)'(\mathbf{A} - \mathbf{B})\mathbf{A}^{-1/2}\boldsymbol{z} = \boldsymbol{z}'\mathbf{A}^{-1/2}(\mathbf{A} - \mathbf{B})\mathbf{A}^{-1/2}\boldsymbol{z} = \boldsymbol{z}'(\mathbf{I} - \mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2})\boldsymbol{z} > 0$. This means that $\mathbf{I} - \mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}$ is positive definite, or, by the result of (a), $\left(\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}\right)^{-1} - \mathbf{I}$ is positive definite. Therefore, for any $\boldsymbol{q} \in \mathbb{R}^n, \boldsymbol{q} \neq \mathbf{0}$, $\boldsymbol{q}'\left(\left(\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}\right)^{-1} - \mathbf{I}\right)\boldsymbol{q} > 0$. Notice that since $\mathbf{A}$ and B are symmetric positive definite matrices, thus their inverse exists, $\left(\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}\right)^{-1} = \mathbf{A}^{1/2}\mathbf{B}^{-1}\mathbf{A}^{1/2}$.

Therefore, for any $\boldsymbol{q} \in \mathbb{R}^n, \boldsymbol{q} \neq \mathbf{0}$, $\boldsymbol{q}' \left(\mathbf{A}^{1/2}\mathbf{B}^{-1}\mathbf{A}^{1/2} - \mathbf{I}\right) \boldsymbol{q} > 0$. Since this last inequality is valid for any $\boldsymbol{q} \in \mathbb{R}^n, \boldsymbol{q} \neq \mathbf{0}$, it must hold for $\boldsymbol{q} = \mathbf{A}^{-1/2}\boldsymbol{z}$, where $\boldsymbol{z} \in \mathbb{R}^n$, $\boldsymbol{z} \neq \mathbf{0}$. Thus, $\boldsymbol{q}' \left(\mathbf{A}^{1/2}\mathbf{B}^{-1}\mathbf{A}^{1/2} - \mathbf{I}\right) \boldsymbol{q} = \boldsymbol{z}'\mathbf{A}^{-1/2} \left(\mathbf{A}^{1/2}\mathbf{B}^{-1}\mathbf{A}^{1/2} - \mathbf{I}\right) \mathbf{A}^{-1/2}\boldsymbol{z} = \boldsymbol{z}' \left(\mathbf{B}^{-1} - \mathbf{A}^{-1}\right) \boldsymbol{z} > 0$, so $\mathbf{B}^{-1} - \mathbf{A}^{-1}$ is positive definite.

Second, we need to show that $\mathbf{B}^{-1} - \mathbf{A}^{-1}$ is positive definite $\Rightarrow$ $\mathbf{A} - \mathbf{B}$ is positive definite. $\mathbf{B}^{-1} - \mathbf{A}^{-1}$ positive definite means that for any $\boldsymbol{p} \in \mathbb{R}^n$, $\boldsymbol{p} \neq \mathbf{0}$, $\boldsymbol{p}'(\mathbf{B}^{-1} - \mathbf{A}^{-1})\boldsymbol{p} > 0$. Let $\boldsymbol{p} = \mathbf{B}^{1/2}\boldsymbol{y}$, where $\boldsymbol{y} \in \mathbb{R}^n, \boldsymbol{y} \neq \mathbf{0}$, then $\boldsymbol{p}'(\mathbf{B}^{-1} - \mathbf{A}^{-1})\boldsymbol{p} = \boldsymbol{y}'\mathbf{B}^{1/2}(\mathbf{B}^{-1} - \mathbf{A}^{-1})\mathbf{B}^{1/2}\boldsymbol{y} = \boldsymbol{y}'(\mathbf{I} - \mathbf{B}^{1/2}\mathbf{A}^{-1}\mathbf{B}^{1/2})\boldsymbol{y} > 0$. This means that $\mathbf{I} - \mathbf{B}^{1/2}\mathbf{A}^{-1}\mathbf{B}^{1/2}$ is positive definite, or, by the result of (a), $\left(\mathbf{B}^{1/2}\mathbf{A}^{-1}\mathbf{B}^{1/2}\right)^{-1} - \mathbf{I}$ is positive definite. Therefore, for any $\boldsymbol{q} \in \mathbb{R}^n, \boldsymbol{q} \neq \mathbf{0}$, $\boldsymbol{q}' \left(\left(\mathbf{B}^{1/2}\mathbf{A}^{-1}\mathbf{B}^{1/2}\right)^{-1} - \mathbf{I}\right) \boldsymbol{q} > 0$. Notice that since $\mathbf{A}$ and $\mathbf{B}$ are symmetric positive definite matrices, thus their inverse exists, $\left(\mathbf{B}^{1/2}\mathbf{A}^{-1}\mathbf{B}^{1/2}\right)^{-1} = \mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$. Therefore, for any $\boldsymbol{q} \in \mathbb{R}^n, \boldsymbol{q} \neq \mathbf{0}$, $\boldsymbol{q}' \left(\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2} - \mathbf{I}\right) \boldsymbol{q} > 0$. Since this last inequality is valid for any $\boldsymbol{q} \in \mathbb{R}^n, \boldsymbol{q} \neq \mathbf{0}$, it must hold for $\boldsymbol{p} = \mathbf{B}^{1/2}\boldsymbol{y}$, where $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{y} \neq \mathbf{0}$. Thus, $\boldsymbol{q}' \left(\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2} - \mathbf{I}\right) \boldsymbol{q} = \boldsymbol{y}'\mathbf{B}^{1/2} \left(\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2} - \mathbf{I}\right) \mathbf{B}^{1/2}\boldsymbol{y} = \boldsymbol{y}' \left(\mathbf{A} - \mathbf{B}\right) \boldsymbol{y} > 0$, so $\mathbf{A} - \mathbf{B}$ is positive definite.

Problem 11. Let $\mathbf{A}$ be a symmetric matrix.

1. Show that the determinant of $\mathbf{A}$ is equal to the product of its eigenvalues.

2. Show that the trace of $\mathbf{A}$ is equal to the sum of its eigenvalues.

3. Show that $\mathbf{A}$ is positive definite (positive semidefinite) if and only if all its eigenvalues are positive (non-negative).

Solution. (1) Claim: The determinant of A is equal to the product of its eigenvalues.

Proof: Theorem: If $\mathbf{A}$ is a real symmetric $n \times n$ matrix, then there exists an orthogonal matrix $\mathbf{P}$ such that $\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{P}'\mathbf{A}\mathbf{P}$ where i) $\mathbf{P}$ is obtained by normalizing a basis of orthogonal eigenvectors of $\mathbf{A}$ and ii) the diagonal elements of $\mathbf{D}$ are the n eigenvalues $\lambda_1, \lambda_2, ...\lambda_n$ of $\mathbf{A}$. Therefore, $\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P} \Rightarrow \mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. $\det(\mathbf{A}) = \det(\mathbf{P}\mathbf{D}\mathbf{P}^{-1}) = \det(\mathbf{P})\det(\mathbf{D})\det(\mathbf{P}^{-1}) = \det(\mathbf{D}) = \prod_{i=1}^n \lambda_i$.

(2) Claim: The trace of A is equal to the sum of its eigenvalues.

Proof: $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{P}\mathbf{D}\mathbf{P}^{-1}) = \mathrm{tr}(\mathbf{P}^{-1}\mathbf{P}\mathbf{D}) = \mathrm{tr}(\mathbf{D}) = \sum_{i=1}^n \lambda_i$.

(3) Claim: A is positive definite (positive semi definite) if and only if all its eigenvalues are positive (non-negative).

First, we want to prove that if $\mathbf{A}$ is positive definite, then all its eigenvalues are positive (necessary part).

$\mathbf{A}$ is positive definite $\Rightarrow$ $\boldsymbol{x}'\mathbf{A}\boldsymbol{x} > 0$ or all vectors $\boldsymbol{x} \neq \mathbf{0}$. We can write: $\boldsymbol{x}'\mathbf{A}\boldsymbol{x} = \boldsymbol{x}'\mathbf{P}'\mathbf{D}\mathbf{P}\boldsymbol{x} = \boldsymbol{v}'\mathbf{D}\boldsymbol{v} = \sum_{i=1}^n \lambda_i v_i^2 > 0$ where $\boldsymbol{v} = \mathbf{P}\boldsymbol{x}$. The result follows as we choose $\boldsymbol{x} = \mathbf{P}^{-1}\boldsymbol{e}_i$, $i = 1, ..., n$, where $\{\boldsymbol{e}_i : i = 1, ..., n\}$ are unit vectors (e.g., $\boldsymbol{e}_1 = (1, 0, ..., 0)$).

Second, we want to prove that if all the eigenvalues of $\mathbf{A}$ are positive, then $\mathbf{A}$ is positive definite (sufficiency part). If all eigenvalues are positive, then $\boldsymbol{x}'\mathbf{A}\boldsymbol{x} = \boldsymbol{x}'\mathbf{P}'\mathbf{D}\mathbf{P}\boldsymbol{x} = \boldsymbol{v}'\mathbf{D}\boldsymbol{v} = \sum_{i=1}^n \lambda_i v_i^2 > 0$ for all vectors $\boldsymbol{x} \neq \mathbf{0}$. So $\mathbf{A}$ is positive definite.

Note: Same idea to show that $\mathbf{A}$ is positive semidefinite.

Problem 12. Let $\mathbf{B}$ be a symmetric and idempotent $n \times n$ matrix: $\mathbf{B}' = \mathbf{B}$ and $\mathbf{B}\mathbf{B} = \mathbf{B}$. Let $\lambda_1, \ldots \lambda_n$ be the eigenvalues of $\mathbf{B}$.

1. Show that the eigenvalues of $\mathbf{B}$ are zeros and/or ones.

2. Show that $\mathrm{rank}\left(\mathbf{B}\right) = \sum_{i=1}^n \lambda_i = \mathrm{tr}\left(\mathbf{B}\right)$.

3. Show that $\mathbf{B}$ is positive semidefinite.

4. Show that $\mathbf{I}_n - \mathbf{B}$ is also symmetric and idempotent.

Solution. (1) Claim: The eigenvalues of B are zeros and/or ones.

Proof: $\mathbf{B}$ is idempotent $\Rightarrow$ $\mathbf{B}\mathbf{B} = \mathbf{B}$. Let λ be an eigenvalue of $\mathbf{B}$ and $\boldsymbol{v}$ be the corresponding eigenvector. Then, $\Rightarrow \lambda\boldsymbol{v} = \mathbf{B}\boldsymbol{v} = \mathbf{B}^2\boldsymbol{v} = \mathbf{B}(\mathbf{B}\boldsymbol{v}) = \mathbf{B}(\lambda\boldsymbol{v}) = \lambda\mathbf{B}\boldsymbol{v} = \lambda^2\boldsymbol{v} \Rightarrow \lambda^2 = \lambda \Rightarrow \lambda = 1$ or 0.

(2) Claim: rank($\mathbf{B}$) $= \sum_{i=1}^{n} \lambda_i =$ tr($\mathbf{B}$)**.**

Proof: We can write $\mathbf{B} = \mathbf{PDP}^{-1}$. Then, rank($\mathbf{B}$) = rank($\mathbf{PDP}^{-1}$) = rank($\mathbf{D}$) since $\mathbf{P}$ and $\mathbf{P}^{-1}$ are invertible matrices. And rank($\mathbf{D}$) = tr($\mathbf{D}$) = the number of linearly independent columns (rows) in $\mathbf{D}$ which is equal to the number of non-zero eigenvalues in $\mathbf{D}$.

(3) Claim: B is positive semidefinite.

Proof: Since all the eigenvalues of $\mathbf{B}$ are either zeros or ones, it is clear that $\mathbf{B}$ is positive semidefinite.

(4) Claim: $\mathbf{I} - \mathbf{B}$ is also symmetric and idempotent.

Proof: Symmetric: $(\mathrm{I} - \mathbf{B})' = \mathbf{I}' - \mathbf{B}' = \mathbf{I} - \mathbf{B}$. Idempotent: $(\mathbf{I} - \mathbf{B})(\mathbf{I} - \mathbf{B}) = \mathbf{I} - \mathbf{B} - \mathbf{B} + \mathbf{B}^2 = \mathbf{I} - 2\mathbf{B} + \mathbf{B} = \mathbf{I} - \mathbf{B}$.