Problems 1

Jakob Sverre Alexandersen
GRA4153 Advanced Statistics

September 4, 2025

Contents

1 Probabilities, random variables

1. A fair die is thrown until a 6 appears. What is the probability that it must be thrown at least k times?

$$
\begin{gathered}
P(\text{at least } k \text{ throws}) = 1 - P(\text{fewer than } k \text{ throws}) \\
P(\text{fewer than } k \text{ throws}) = P(\text{success in first } k-1 \text{ throws}) \\
P(\text{success in first } k-1 \text{ throws}) = 1 - \left(\frac{5}{6}\right)^{k-1}
\end{gathered}
$$

2. For $x \in \mathbb{R}^K$, let $f_\theta(x) := h(x)\exp(\eta(\theta)'T(x) - A(\theta))$ for functions h, η, T, A. When is this a pdf? i.e. when is $f(x) \geq 0$ and such that its integral is equal to one?

Non-negativity: $f_\theta(x) \geq 0$

$$
\begin{gathered}
h(x) \geq 0 \quad \forall x \text{ in the support} \\
\exp(\eta(\theta)'T(x) - A(\theta)) \geq 0
\end{gathered}
$$

Integrating to 1:

This is where the log-partition function $A(\theta)$ plays a crucial role:

$$
\begin{gathered}
\int f_\theta(x)dx = \int h(x)\exp(\eta(\theta)'T(x) - A(\theta))dx = 1 \\
= \exp(-A(\theta)) \int h(x)\exp(\eta(\theta)'T(x))dx = 1 \\
\therefore A(\theta) = \log \int h(x)\exp(\eta(\theta)'T(x))dx
\end{gathered}
$$

3. For each $i = 1, \ldots, K$, let f_i be pdfs (resp.) and define $f(x) := \sum_{i=1}^{K} w_i f_i(x)$ where $w_i \geq 0$ and $\sum_{i=1}^{K} w_i = 1$. Show that f is a probability density (resp. probability mass) function. i.e. show that $f(x) \geq 0$ and its integral / sum is equal to one in the density / mass case respectively

Here we need to verify two conditions:

- **Non-negativity:** Show that $f(x) \geq 0 \quad \forall x$
- **Normalization:** Show that the integral (or sum) equals 1

Non-negativity:

Since each $f_i(x)$ is a valid pdf/pmf, we have $f_i(x) \geq 0 \quad \forall x \wedge (i = 1, \ldots, K)$

Additionally, we have given that $w_i \geq 0 \quad \forall i = 1, \ldots, K$

$$\therefore f(x) = \sum_{i=1}^{K} w_i f_i(x) \geq 0$$

Since we are summing non-negative terms ($w_i \geq 0 \wedge f_i \geq 0$)

Normalization:

Case 1: pdf

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \sum_{i=1}^{K} w_i f_i(x)dx$$

$$= \sum_{i=1}^{K} w_i \int_{-\infty}^{\infty} f_i(x)dx \quad \text{(linearity of integration)}$$

$$= \sum_{i=1}^{K} w_i \times 1 \quad \text{(since each } f_i \text{ is a valid pdf)}$$

$$= \sum_{i=1}^{K} w_i$$

$$= 1 \quad \text{(given the constraint } \sum_{i=1}^{K} w_i = 1\text{)}$$

Case 2: pmf

$$\sum_x f(x) = \sum_x \sum_{i=1}^{K} w_i f_i(x)$$

$$= \sum_{i=1}^{K} w_i \sum_x f_i(x) \quad \text{(linearity of summation)}$$

$$= \sum_{i=1}^{K} w_i \times 1 \quad \text{(since each } f_i \text{ is a valid pmf)}$$

$$= \sum_{i=1}^{K} w_i$$

$$= 1 \quad \text{(given constraint)}$$

4. Let X be a Poisson r.v. with mass function $f(x) = \lambda^x \exp(-\lambda)/x!, \quad x = 0, 1, \ldots$ for $\lambda > 0$. Find the probability that X is odd

$$\exp(\lambda) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = 1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \ldots + \frac{\lambda^n}{n!}$$
$$\exp(-\lambda) = \sum_{x=0}^{\infty} \frac{-\lambda^x}{x!} = 1 + \frac{-\lambda}{1!} + \frac{-\lambda^2}{2!} + \ldots + \frac{-\lambda^n}{n!}$$
$$P(X \text{ is odd}) = \sum_{x \text{ odd}} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x \text{ odd}} \frac{\lambda^x}{x!}$$

$$e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \quad \text{(sum of all terms)}$$
$$e^{-\lambda} = \sum_{x=0}^{\infty} \frac{(-1)^x \lambda^x}{x!} \quad \text{(alternating signs)}$$

$$e^{\lambda} + e^{-\lambda} = 2 \sum_{x \text{ even}} \frac{\lambda^x}{x!} \quad \text{(even terms don't cancel)}$$

$$e^{\lambda} - e^{-\lambda} = 2 \sum_{x \text{ odd}} \frac{\lambda^x}{x!} \quad \text{(odd terms don't cancel)}$$

$$\rightarrow \sum_{x \text{ odd}} \frac{\lambda^x}{x!} = \frac{e^{\lambda} - e^{-\lambda}}{2}$$

$$P(X \text{ is odd}) = e^{-\lambda} \cdot \frac{e^{\lambda} - e^{-\lambda}}{2} = \frac{1 - e^{-2\lambda}}{2}$$

5. Prove that $F(x) := (1+\exp(-x))^{-1} \quad x \in \mathbb{R}$ is a CDF

Recall that a CDF must satisfy these requirements:

- **Monotonicity:** F is non-decreasing (i.e. $x_1 \leq x_2 \to F(x_1) \leq F(x_2)$)
- **Right-continuity:** F is right-continuous at every point
- **Limit conditions:**
  - $\lim_{x\to-\infty} F(x) = 0$
  - $\lim_{x\to\infty} F(x) = 1$

Limit conditions: It is trivial that the function satisfies these two conditions.

Monotonicity: We prove that $F'(x) \geq 0$:

$$F'(x) = \frac{\exp(-x)}{\big(1+\exp(-x)\big)^2} \quad \forall x \in \mathbb{R}$$

Right-continuity:

Since $F(x)$ is continuous everywhere (as a composition of continuous functions), it is automatically right-continuous

$\therefore F(x)$ is a CDF $\square$

6. Show that any CDF F, i.e. $F(x) := P(X \leq x)$, can have at most a countable number of discontinuities

The key here is to use the monotonicity of CDFs combined with the fact that rational numbers are countable.

For any CDF F, discontinuities can only by "jump" discontinuities due to monotonicity. At each discontinuity point x_0 we have:

- Left limit: $F(x_0^-) = \lim_{x \to x_0^-} F(x)$ exists
- Right limit: $F(x_0^+) = \lim_{x \to x_0^+} F(x) = F(x_0) \quad$ (right-continuity)
- Jump size: $F(x_0) - F(x_0^-) > 0$

Associate each discontinuity with a rational numebr:

- Let D be the set of discontinuity points
- For each $x \in D$, the jump size is $F(x) - F(x^-) > 0$
- Between any two consecutive jumps $F(x^-)$ and $F(x)$, there exists a rational number
- Since F is monotonic, these rational intervals are disjoint

Rationals are countable:

- Each discontinuity corresponds to a unique rational number in $(F(x^-), F(x)]$
- Since $\mathbb{Q} \cap [0, 1]$ is countable, and all these rationals are distinct
- Therefore D is at most countable

2 Expectations

1. Show that $\mathbb{E}[\alpha] = \alpha$ for any non-random α

By the definition of the expected value (and the fact that all pdfs integrate to one):

$$\mathbb{E}[c] = \int_{-\infty}^{\infty} cf(x)\ dx = c\int_{-\infty}^{\infty} f(x)\ dx = c \cdot 1 = c$$

2. Let X be the sum of two rolls of a fair die. What is the mean and variance of X?

X itself is defined as $X = X_1 + X_2$, where $X_1, X_2 \sim \text{Uniform}\{1, \dots, 6\}$, independent.

This means that:

$$\mathbb{E}[X] = \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2]$$

Since the die are identical, $\mathbb{E}[X_1] = \mathbb{E}[X_2]$, and thus:

$$\mathbb{E}[X_1] = \sum_{i=1}^{6} P(x = i) = \frac{1}{6}\sum_{i=1}^{6} i = \frac{7}{2}$$

$$\mathbb{E}[X] = \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = \frac{7}{2} + \frac{7}{2} = 7$$

Since the die are independent, we only need to compute the one variance:

$$\begin{gathered}\text{Var}(X) = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2 \\ \mathbb{E}[X_1^2] = \frac{1}{6}\sum_{i=1}^{6} i^2 = \frac{91}{6} \\ \text{Variance of one die: } \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}\end{gathered}$$

$$\text{Var}(X) = \text{Var}(X_1) + \text{Var}(X_2) = 2 \cdot \frac{35}{12} = \frac{35}{6} = 5.8\bar{3}$$

3. X is uniformly distributed on $[a,b]$ if its density is $f(x) = \frac{1}{b-a}$. Compute the mean and variance of X. Recall that $\mathbb{E}[x] = \int_{-\infty}^{\infty} xf(x)\, dx$

$$\mathbb{E}[X] = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{1}{b-a} \cdot \int_a^b x dx$$

$$= \frac{1}{b-a} \cdot \frac{x^2}{2} = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b^2-a^2}{2(b-a)}$$

$$= \frac{b+a}{2}$$

Nice. Also recall that $\mathrm{Var}(X) = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2$:

$$\mathbb{E}[X^2] = \int_a^b x^2 \cdot \frac{1}{b-a} dx = \frac{1}{b-a} \cdot \int_a^b x^2 dx$$

$$= \frac{1}{b-a} \cdot \frac{x^3}{3} = \frac{x^3}{3(b-a)}\Big|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}$$

Thus:

$$\mathrm{Var}[X] = \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2 = \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4}$$

$$= \frac{(b-a)^2}{12}$$

4. Calculate the mean of $X \sim t(v)$. Are restrictions on v required for the mean to exist?

First we should define the student's t-distribution:

$$f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}\Gamma\left(\frac{v}{2}\right)}\left(1 + \frac{t^2}{v}\right)^{\frac{-(v+1)}{2}}$$

By symmetry the integrand $xf(x)$ is an odd function, so if the expectation integral converges, then

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)\ dx = 0$$

Existence: for large $|x|$, the density behaves like a constant times $|x|^{-(v+1)}$. Hence:

$$\mathbb{E}[|X|] \asymp \int_{1}^{\infty} x \cdot x^{-(v+1)}\ dx = \int_{1}^{\infty} x^{-v}\ dx$$

Which converges iff $v > 1$

5. Prove Proposition 2.6 for the discrete case.

Prop. 6:

Let X be a random variable, α a constant and g_1 and g_2 such that $\mathbb{E}g_1(X)$ and $\mathbb{E}g_2(X)$ exist. Then

$$i) \quad \mathbb{E}[\alpha g_1(X)] = \alpha \mathbb{E}g_1(X) \text{ and } \mathbb{E}[g_1(X) + g_2(X)] = \mathbb{E}g_1(X) + \mathbb{E}g_2(X);$$
$$ii) \quad \text{If } g_1(x) > 0 \quad \forall x \text{ with } f(x) > 0, \text{ then } \mathbb{E}g_1(X) \geq 0$$

Let X be a discrete r.v. with pmf $P(X = x)$, where x ranges over the support of X

Part i): linearity of expectation

For scalar multiplication:

$$\mathbb{E}[\alpha g_1(X)] = \sum_{x \in \mathcal{X}} \alpha g_1(x) P(X = x) = \alpha \sum_{x \in \mathcal{X}} g_1 P(X = x) = \alpha \mathbb{E}[g_1(X)]$$

For addition:

$$\mathbb{E}[g_1(X) + g_2(X)] = \sum_{x \in \mathcal{X}} [g_1(x) + g_2(x)] P(X = x)$$

$$= \sum_{x \in \mathcal{X}} g_1 P(X = x) + \sum_{x \in \mathcal{X}} g_2 P(X = x)$$

$$= \mathbb{E}[g_1(X)] + \mathbb{E}[g_2(X)]$$

Part ii): non-negative property:

If $g_1(x) \geq 0 \quad \forall x$ with $P(X = x)$, then:

$$\mathbb{E}[g_q(X)] = \sum_{x \in \mathcal{X}} g_1 P(X = x)$$

Since each term $g_1(x) P(X = x) \geq 0$ (because $g_1(x) \geq 0$ and $P(X = x) \geq 0$), we have:

$$\mathbb{E}[g_1(X)] = \sum_{x \in \mathcal{X}} g_1(x) P(X = x) \geq 0$$

6. Prove Lemma 2.1:

It states that *If* $\mathrm{Var}(X)$ *exists, then for any constants* a, b

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$$

Recall that $\mathrm{Var}(X) = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2$. Therefore,

$$\begin{aligned}\mathbb{E}[(aX+b)^2] - \big(\mathbb{E}[(aX+b)]\big)^2 &= \mathbb{E}\big[(aX+b) - \mathbb{E}[(aX+b)]^2\big] \\ &= \mathbb{E}\big[(aX - a\mathbb{E}[X])^2\big] = a^2\mathbb{E}\big[(X-\mathbb{E}[X])^2\big] \\ &= a^2\mathrm{Var}(X)\end{aligned}$$

$\square$

7. Prove Lemma 2.2:

It states that [Properties of indicators]: *If* A, B *are events and* X *an r.v.:*

$$\begin{aligned}(i) &\quad \mathbb{1}_A\mathbb{1}_B = \mathbb{1}_{A\cap B} \\ (ii) &\quad P(X\in A) = \mathbb{E}[\mathbb{1}_A(X)] \\ (iii) &\quad P(X\in A)[1-P(X\in A)] = \mathrm{Var}(\mathbb{1}_A(X))\end{aligned}$$

For an event E, its indicator $\mathbb{1}_E$ takes value 1 on E and 0 on E^c. If X is an r.v. and A is a measurable set in the state space of X, then $\mathbb{1}_A(X)$ means the indicator of the event $\{\omega : X(\omega)\in A\}$, i.e. $\mathbb{1}_{X\in A}$

$$(i) \quad \mathbb{1}_A\mathbb{1}_B = \mathbb{1}_{A\cap B}$$

Pointwise check: for any ω, the left side is 1 iff. $\omega\in A \;\wedge\; \omega\in B$, otherwise it is 0. That is exactly the indicator of $A\cap B$ $\square$

$$(ii) \quad P(X\in A) = \mathbb{E}[\mathbb{1}_A(X)]$$

Let $E = \{\omega : X(\omega)\in A\} = X^{-1}(A)$. Then $\mathbb{1}_A(X) = \mathbb{1}_E$

For any event E, $\mathbb{E}[\mathbb{1}_E] = \int \mathbb{1}_E \, dP = P(E)$. Hence $\mathbb{E}[\mathbb{1}_A(X)] = P(X\in A)$ $\square$

$$(iii) \quad P(X\in A)[1-P(X\in A)] = \mathrm{Var}(\mathbb{1}_A(X))$$

Set $Y = \mathbb{1}_A(X)$. Then $Y\in\{0,1\}$ and $\mathbb{E}[Y] = P(X\in A) =: p$

Since $Y^2 = Y$, $\mathrm{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = p - p^2 = p(1-p)$

$\therefore \mathrm{Var}\big(\mathbb{1}_A(X)\big) = P(X\in A)\big[1-P(X\in A)\big]$ $\square$

8. Let X and Y be r.v.s with $\mathbb{E}|X| < \infty$, $\mathbb{E}|Y| < \infty$ and let $X \wedge Y := \min\{X, Y\}$ and $X \vee Y := \max\{X, Y\}$. Show that $\mathbb{E}[X \vee Y] = \mathbb{E}[X] + \mathbb{E}[Y] - \mathbb{E}[X \wedge Y]$ [Hint: What is $(X \vee Y) + (X \wedge Y)$?]

Consider what happens when we add the max and min of two numbers: $\min\{a, b\} + \max\{a, b\} = a + b$

$\therefore (X \vee Y) + (X \wedge Y) = X + Y$

Since $(X \vee Y) + (X \wedge Y) = X + Y$, we can carry expectation operations:

$$\mathbb{E}[(X \vee Y) + (X \wedge Y)] = \mathbb{E}[X + Y]$$

We can use linearity of expectation since $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$:

$$\mathbb{E}[X \vee Y] + \mathbb{E}[X \wedge Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

$$= \mathbb{E}[X \vee Y] = \mathbb{E}[X] + \mathbb{E}[Y] - \mathbb{E}[X \wedge Y]$$

$\square$

3 Conditioning & independence

1. If P is a probability and B an event with $P(B) > 0$, show that $P(\cdot|B)$ is also a probability

$P(\cdot|B)$ must satisfy the three axioms of probability.

Definition: For an event A, we have $P(A|B) = \frac{P(A\cap B)}{P(B)}$ where $P(B) > 0$

Axiom 1: Non-negativity: For any event A, we need $P(A|B) \geq 0$.

Since P is a probability measure:

$$P(A \cap B) \geq 0 \quad \text{(by non-negativity of } P\text{)}$$
$$P(B) > 0 \quad \text{(given condition)}$$
$$\therefore P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0$$

Axiom 2: Normalization: we need $P(\Omega|B) = 1$, where Ω is the sample space.

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1 \quad (\text{since } \Omega \cap B = B)$$

Axiom 3: Countable additivity: For countably many disjoints events $A_1, \ldots, A_n$, we need:

$$P\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) = \sum_{i=1}^{\infty} P(A_i|B)$$

$$P\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) = \frac{P\big((\bigcup_{i=1}^{\infty} A_i) \cap B\big)}{P(B)}$$

$$= \frac{P\big(\bigcup_{i=1}^{\infty} (A_i \cap B)\big)}{P(B)} \quad \text{(using the distributive property of intersection over union)}$$

$$= \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i|B)$$

Since the A_i are disjoint, the events $A_i \cap B$ are also disjoint. Therefore, by cointable additivity of P:

Since $P(\cdot|B)$ satisfies all three axioms of probability, it is indeed a probability measure $\square$

2. If $P(B \cap C) > 0$ show that $P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C)$

We know that $P(A \cap B) = P(A|B)P(B)$

Let $B_c = B \cap C \quad \therefore P(A \cap B_c) = P(A|B \cap C)P(B \cap C)$

By the definition: $P(B \cap C) = P(B|C)P(C)$

Subsitute back in: $P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C)$ $\square$

3. Prove Lemma 2.3

It states that *If* $X = (X_1, \ldots, X_K)$ *is a random vector and* $\text{Var}(X)$ *exists then for any (constant) vector* $\vec{b} \in \mathbb{R}^K$ *and any (constant) matrix* $A \in \mathbb{R}^{m \times K}$*,*

$$\text{Var}(AX + b) = A\text{Var}(X)A'$$

Fix dimensions: $X \in \mathbb{R}^K$ and $A \in \mathbb{R}^{m \times K} \to AX \in \mathbb{R}^m$

Let $Y = AX + b$, where $X \in \mathbb{R}^K, A \in \mathbb{R}^{m \times K}$, and $b \in \mathbb{R}^m$ (note that the dimensions of b must match AX)

Thus,

$$\begin{aligned}
\mathbb{E}[YY'] &= \mathbb{E}[(AX + b)(AX + b)'] \\
&= \mathbb{E}[AXX'A' + AX'b + bX'A' + bb'] \\
&= A\mathbb{E}[XX']A' + A\mathbb{E}[X]b' + b\mathbb{E}[X]'A' + bb'
\end{aligned}$$

Also,

$$\begin{aligned}
\mathbb{E}[Y]\mathbb{E}[Y]' &= (A\mathbb{E}[X] + b)(A\mathbb{E}[X] + b)' \\
&= A\mathbb{E}[X]\mathbb{E}[X]'A' + A\mathbb{E}[X]b' + b\mathbb{E}[X]'A' + bb'
\end{aligned}$$

$$\begin{aligned}
\therefore \text{Var}(AX + b) &= \mathbb{E}[Y]\mathbb{E}[Y]' \\
&= A(\mathbb{E}[XX'] - \mathbb{E}[X]\mathbb{E}[X]')A' \\
&= A\text{Var}(X)A'
\end{aligned}$$

4. Prove Lemma 2.4

Lemma 2.4: *If X, Z and Y are random vectors in* $\mathbb{R}^K, \mathbb{R}^K$ *and* $\mathbb{R}^L$*, respectively,* $a \in \mathbb{R}^K, b \in \mathbb{R}^L$ *are constant vectors and A and B are constant matrices with K and L columns respectively,*

$$\text{Cov}(X + Z, Y) = \text{Cov}(X, Y) + \text{Cov}(Z, Y)$$
$$\textit{and}$$
$$\text{Cov}(AX + a, BY + b) = A\text{Cov}(X, Y)B'$$

Recall that the matrix covariance of random vectors $X \in \mathbb{R}^K$ and $Y \in \mathbb{R}^L$ is

$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)']$

Additivity in the first argument:

Let $\mu_X = \mathbb{E}X, \quad \mu_Z = \mathbb{E}Z, \quad \mu_Y = \mathbb{E}Y$, then:

$$\begin{gathered} \text{Cov}(X + Z, Y) \\ = \mathbb{E}[(X + Z - \mu_X - \mu_Z)(Y - \mu_Y)'] \\ = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)' + (Z - \mu_Z)(Y - \mu_Y)'] \\ = \text{Cov}(X, Y) + \text{Cov}(Z, Y) \end{gathered}$$

$\square$

Affine equivariance:

Let $U = AX + a$ and $V = BY + b$. Then $\mathbb{E}U = A\mathbb{E}X + a$ and $\mathbb{E}V = B\mathbb{E}Y + b$, so

$$\begin{gathered} \text{Cov}(U, V) \\ = \mathbb{E}[(U - \mathbb{E}U)(V - \mathbb{E}V)'] \\ = \mathbb{E}[(AX + a - (A\mathbb{E}X + a))(BY + b - (B\mathbb{E}Y + b))'] \\ = \mathbb{E}[(A(X - \mathbb{E}X))(B(Y - \mathbb{E}Y))'] \\ = \mathbb{E}[A(X - \mathbb{E}X)(Y - \mathbb{E}Y)'B'] \\ = A\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)']B' \\ = A\text{Cov}(X, Y)B' \end{gathered}$$

$\square$

5. Prove Lemma 2.5

If X, Z and Y are r.vs. then

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$$

If X and Y are random vectors of the same dimension then

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Y)'$$

Key identity: $\mathrm{Var}(Z) = \mathrm{Cov}(Z,Z)$, and for any r.v./vectors U, V, W,

$$\begin{aligned}\mathrm{Cov}(U+V,W) &= \mathrm{Cov}(U,W) + \mathrm{Cov}(V,W)\\ \mathrm{Cov}(U,V+W) &= \mathrm{Cov}(U,V) + \mathrm{Cov}(U,W)\end{aligned}$$

Scalar case (real-valued X, Y):

$$\begin{gathered}\mathrm{Var}(X+Y) = \mathrm{Cov}(X+Y, X+Y)\\ = \mathrm{Cov}(X,X) + \mathrm{Cov}(X,Y) + \mathrm{Cov}(Y,X) + \mathrm{Cov}(Y,Y)\\ = \mathrm{Var}(X) + \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Y) + \mathrm{Var}(Y)\\ = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)\end{gathered}$$

Vector case ($X, Y \in \mathbb{R}^d$):

Use $\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)']$ (a $d \times d$ matrix) so $\mathrm{Cov}(Y,X) = \mathrm{Cov}(X,Y)'$:

$$\begin{gathered}\mathrm{Var}(X+Y) = \mathrm{Cov}(X+Y, X+Y)\\ = \mathrm{Cov}(X,X) + \mathrm{Cov}(X,Y) + \mathrm{Cov}(Y,X) + \mathrm{Cov}(Y,Y)\\ = \mathrm{Var}(X) + \mathrm{Var} + \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Y)'\end{gathered}$$

□

6. Prove Corollary 2.1

If X and Y are independent, then $\mathrm{Cov} = 0$

$$\mathrm{Cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Using conditional expectation,

$$\mathbb{E}[XY] = \mathbb{E}\big[\mathbb{E}[XY|X]\big] = \mathbb{E}\big[X\mathbb{E}[Y|X]\big]$$

If X and Y are independent, then $\mathbb{E}[Y|X] = \mathbb{E}[Y]$. Hence $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, and therefore $\mathrm{Cov}(X, Y) = 0$

□

7. Prove Proposition 2.10

Let X and Y be random vectors, α a constant and g_1 and g_2 such that $\mathbb{E}g_1(X)$ and $\mathbb{E}g_2(X)$ exist. Then

$$(i) \quad \mathbb{E}[\alpha g_1(X)|Y] = \alpha\mathbb{E}[g_1(X)|Y] \textit{ and } \mathbb{E}[g_1(X) + g_2(X)|Y] = \mathbb{E}[g_1(X)|Y] + \mathbb{E}[g_2(X)|Y];$$
$$(ii) \quad \textit{If } g_1 \geq 0 \quad \forall x \textit{ with } f(x|y) > 0, \textit{ then } \mathbb{E}[g_1(X)|Y] \geq 0$$

Let $Z_1 := g_1(X)$ and $Z_2 := g_2(X)$, and let $\mathcal{G} := \sigma(Y)$. Recall the defining property of conditional expectation: For any integrable Z, a version of $\mathbb{E}[Z|\mathcal{G}]$ is the $\mathcal{G}$-measurable W s.t. for every $A \in \mathcal{G}$, $\mathbb{E}[Z\mathbf{1}_A] = \mathbb{E}[W\mathbf{1}_A]$

Such W is unique a.s.

(i) Linearity:

Scalar multiple: for any $A \in \mathcal{G}$,

$$\mathbb{E}[(\alpha Z_1)\mathbf{1}_A] = \alpha\mathbb{E}[Z_1\mathbf{1}_A] = \alpha\mathbb{E}\big[\mathbb{E}[Z_1|\mathcal{G}]\mathbf{1}_A\big] = \mathbb{E}\big[\alpha\mathbb{E}[Z_1|\mathcal{G}]\mathbf{1}_A\big]$$

Both $\mathbb{E}[\alpha Z_1|\mathcal{G}]$ and $\alpha\mathbb{E}[Z_1]\mathcal{G}$ are $\mathcal{G}$-measurable and satisfy the same defining identity, hence

$$\mathbb{E}[\alpha Z_1|\mathcal{G}] = \alpha\mathbb{E}[Z_1|\mathcal{G}] \quad \text{a.s.}$$

Replacing Z_1 by $g_1(X)$ and $\mathcal{G}$ by $\sigma(Y)$ gives $\mathbb{E}[\alpha g_1(X)|Y] = \alpha\mathbb{E}[g_1(X)|Y]$ a.s.

Sum: For any $A \in \mathcal{G}$,

$$\mathbb{E}[(Z_1 + Z_2)\mathbf{1}_A] = \mathbb{E}[Z_1\mathbf{1}_A] + \mathbb{E}[Z_2\mathbf{1}_A] = \mathbb{E}\big[\mathbb{E}[Z_1|\mathcal{G}]\mathbf{1}_A\big] + \mathbb{E}\big[\mathbb{E}[Z_2|\mathcal{G}]\mathbf{1}_A\big] = \mathbb{E}\big[(\mathbb{E}[Z_1|\mathcal{G}] + \mathbb{E}[Z_2|\mathcal{G}])\mathbf{1}_A\big]$$

By the same uniqueness argument,

$$\mathbb{E}[Z_1 + Z_2|\mathcal{G}] = \mathbb{E}[Z_1|\mathcal{G}] + \mathbb{E}[Z_2|\mathcal{G}] \quad \text{a.s.}$$

Substituting $Z_i = g_i(X)$ and $\mathcal{G} = \sigma(Y)$ gives the desired result □

(ii) Positivity:

Assume $g_1(X) \geq 0$ a.s. Let $W := \mathbb{E}[g_1(X)|\mathcal{G}]$. For any $A \in \mathcal{G}$,

$$\mathbb{E}[W\mathbf{1}_A] = \mathbb{E}[g_1(X)\mathbf{1}_A] \geq 0$$

If $\mathbb{P}(W < 0) > 0$, take $A = \{W < 0\} \in \mathcal{G}$; then $\mathbb{E}[W\mathbf{1}_A] < 0$, a contradiction. Hence $\mathbb{P}(W \geq 0) = 1$, i.e.

$$\mathbb{E}[g_1(X)|Y] \geq 0 \quad \text{a.s.}$$

$\square$