Problem Set Week 6 Solutions

ETHZ Math Olympiad Club

24 March 2025

1 Problem (unknown)

We consider a game where two indistinguishable envelopes are presented to a player:

- One envelope contains an amount $\alpha \in \mathbb{R}_{>0}$.
- The other envelope contains 2α.

The game proceeds as follows:

1. The player randomly selects one envelope (with equal probability).
2. The player observes the content x of the selected envelope (without knowing α).
3. The player must decide whether to:
   - Keep the current envelope, or
   - Switch to the other envelope (this decision is irrevocable).

Although the game is played once, the player's objective is still to maximize their *expected gain*. Assuming access to *randomness*, how can they do better than always keeping the first envelope?

Answer:

As outlined in the problem, we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω consists of two possible configurations of envelopes: $\Omega = \{(\alpha, 2\alpha), (2\alpha, \alpha)\}$. Since Ω has only two elements, the only choice for the sigma-algebra (other than the trivial one) is its power set, $\mathcal{F} = \mathscr{P}(\Omega)$. The player selects an envelope at random with equal probability, implying that the probability measure $\mathbb{P}$ follows Laplace's model. Thus, for any $A \in \mathcal{F}$:

$$\mathbb{P}(A) := \frac{|A|}{|\Omega|} = \frac{|A|}{2}.$$

In this framework, the "naive" strategies—always keeping the selected envelope or always switching—can be analyzed by computing the expected value of the first coordinate projection $\pi_1 : \Omega \to \mathbb{R}$ (or equivalently, the second coordinate π_2). These functions are obviously measurable and positive, and we compute:

$$\mathbb{E}(\pi_i) = \alpha \mathbb{P}_{\pi_i}(\{\alpha\}) + 2\alpha \mathbb{P}_{\pi_i}(\{2\alpha\}) = \alpha \frac{1}{2} + 2\alpha \frac{1}{2} = \frac{3}{2}\alpha.$$

However, since we assume access to randomness, we can improve upon these deterministic strategies by incorporating some randomization. The key observation is that while the player

does not know α, i.e., they know only "partially" Ω (having access to Ω only through π_1; opening one of the envelopes only gives them a value x and the other one may be $2x$ or $\frac{x}{2}$). They still know that one value is strictly greater than the other, i.e., $\alpha < 2\alpha$.
To exploit this without knowing Ω, the player, after having selected one envelope and having seen its content, introduces some randomization. Say, they will generate[1] a number between $[0,1]$ and *make a decision* with this number. Formally, we introduce an auxiliary probability space $(\Sigma, \mathcal{A}, \mu)$ and let $U : \Sigma \to [0,1]$ be a uniform random variable, i.e., $U \sim \text{Unif}\left([0,1]\right)$[2], that is, for any $a, b \in [0,1]$ with $a \leq b$ and any non-empty interval $I \subset [0,1]$ with $\inf I = a$ and $\sup I = b$ then $\mu_U(I) = b - a$. Adjoining this new probability space to the preceding one, we obtain a product space (which is a probability space):

$$\left(\Omega \times \Sigma, \mathcal{F} \otimes \mathcal{A}, \mathbb{P} \otimes \mu\right).$$

Choosing a $(w, \xi) \in \Omega \times \Sigma$, and looking at $\pi_1(w)$, we must now make a decision about switching or not with a deterministic choice involving the pair $\left(\pi_1(w), U(\xi)\right)$. That is, we have a measurable function $f : \mathbb{R}_{>0} \times [0,1] \to \{0,1\}$ that takes $\left(\pi_1(w), U(\xi)\right)$ and if $f\left(\left(\pi_1(w), U(\xi)\right)\right) = 0$, we keep, i.e., we take $\pi_1(w)$, and if $f\left(\left(\pi_1(w), U(\xi)\right)\right)$ is 1, we switch, i.e., we take $\pi_2(w)$. Formally, the gain with respect to this choice of f is $G_f : \Omega \times \Sigma \to \mathbb{R}$ where:

$$G_f\left((w,\xi)\right) = \pi_1(w) \cdot \mathbb{1}_{\{0\}}\left(f\left(\left(\pi_1(w), U(\xi)\right)\right)\right) + \pi_2(w) \cdot \mathbb{1}_{\{1\}}\left(f\left(\left(\pi_1(w), U(\xi)\right)\right)\right),$$

which is measurable as long as f is measurable. Then, with probability $\frac{1}{2}$, the first envelope has amount α, and we decide to keep it with a probability $\mu_{f(\alpha,U)}\left(\{0\}\right)$ and we switch with probability $\mu_{f(\alpha,U)}\left(\{1\}\right)$. With probability $\frac{1}{2}$, the first envelope has amount 2α, and we decide to keep it with probability $\mu_{f(2\alpha,U)}\left(\{0\}\right)$ and we switch with probability $\mu_{f(2\alpha,U)}\left(\{1\}\right)$. The expected gain in this new probability space is given by:

$$\begin{gathered}\mathbb{E}\left(G_f\right) = \\ \alpha\left(\mathbb{P}_{\pi_1}\left(\{\alpha\}\right)\mu_{f(\alpha,U)}\left(\{0\}\right) + \mathbb{P}_{\pi_1}\left(\{2\alpha\}\right)\mu_{f(2\alpha,U)}\left(\{1\}\right)\right) \\ + \\ 2\alpha\left(\mathbb{P}_{\pi_1}\left(\{2\alpha\}\right)\mu_{f(2\alpha,U)}\left(\{0\}\right) + \mathbb{P}_{\pi_1}\left(\{\alpha\}\right)\mu_{f(\alpha,U)}\left(\{1\}\right)\right) \\ = \frac{1}{2}\alpha\left(\mu_{f(\alpha,U)}\left(\{0\}\right) + \mu_{f(2\alpha,U)}\left(\{1\}\right)\right) + \alpha\left(\mu_{f(2\alpha,U)}\left(\{0\}\right) + \mu_{f(\alpha,U)}\left(\{1\}\right)\right).\end{gathered}$$

This formulation may appear abstract, and the machinery might seem overdeveloped, but the goal is to introduce the object naturally and provide a philosophical motivation for what follows.

We observe an asymmetry in this linear combination of α and 2α, which suggests an opportunity to exploit it in order to increase the expected gain beyond $\frac{3}{2}\alpha$. A natural way to introduce randomness into our decision-making process (that is, a natural choice of f) is by defining a threshold: if $U(\xi)$ is below a certain threshold, we keep the first envelope; otherwise, we switch. However, since we have access to the amount in the first envelope, we can dynamically adjust this threshold to take advantage of the inequality $\alpha < 2\alpha$ and the

[1] Practically, this could involve sampling from a physical source (e.g., thermal noise) or a deterministic pseudorandom number generator (PRNG) seeded by an unpredictable value (e.g., system clock nanoseconds); see Hardware random number generator for more information.

[2] We can construct $(\Sigma, \mathcal{A}, \mu)$ as $\left([0,1], \mathcal{B}\left([0,1]\right), \lambda|_{\mathcal{B}([0,1])}\right)$, where $\mathcal{B}\left([0,1]\right)$ is the Borel sigma-algebra and λ is the Lebesgue measure on $\mathbb{R}$. U can then be defined as the identity function, i.e., $U(x) = x$ for $x \in [0,1]$ and it is obviously a random variable with the uniform law on $[0,1]$.

asymmetry in the expected gain formula.

To formalize this idea, we introduce a measurable function $g : \mathbb{R}_{>0} \to [0,1]$ and define a decision function $f_g : \mathbb{R}_{>0} \times [0,1] \to \{0,1\}$ by

$$f_g\left(x,y\right) = \mathbb{1}_{[0,g(x)]}\left(y\right),$$

which means we keep the first envelope if the generated number $U\left(\xi\right)$ satisfies $U\left(\xi\right) \leq g\left(\pi_1\left(\omega\right)\right)$, and we switch otherwise.

For this specific choice of f_g, the probabilities of keeping and switching the envelope are as follows: with probability $\frac{1}{2}$ the first envelope has amount α, and we keep it if the randomly generated number in $[0,1]$ is below $g\left(\alpha\right)$ that is we keep it with probability:

$$\mu_{f_g(\alpha,U)}\left(\{0\}\right) = \mu\left(\{\xi \in \Sigma \,|\, U\left(\xi\right) \in [0, g\left(\alpha\right)]\}\right) = g\left(\alpha\right)$$

and we switch with probability:

$$\mu_{f_g(\alpha,U)}\left(\{1\}\right) = \mu\left(\{\xi \in \Sigma \,|\, U\left(\xi\right) \in]g\left(\alpha\right), 1]\}\right) = 1 - g\left(\alpha\right).$$

Similarly, with probability $\frac{1}{2}$ the first envelope has amount 2α, and we keep it with probability:

$$\mu_{f_g(2\alpha,U)}\left(\{0\}\right) = \mu\left(\{\xi \in \Sigma \,|\, U\left(\xi\right) \in [0, g\left(2\alpha\right)]\}\right) = g\left(2\alpha\right)$$

and we switch with probability:

$$\mu_{f_g(2\alpha,U)}\left(\{1\}\right) = \mu\left(\{\xi \in \Sigma \,|\, U\left(\xi\right) \in]g\left(2\alpha\right), 1]\}\right) = 1 - g\left(2\alpha\right).$$

Thus, the expected gain under this strategy is:

$$\begin{aligned}
\mathbb{E}\left(G_{f_g}\right) &= \frac{1}{2}\alpha\left(\mu_{f_g(\alpha,U)}\left(\{0\}\right) + \mu_{f_g(2\alpha,U)}\left(\{1\}\right)\right) + \alpha\left(\mu_{f_g(2\alpha,U)}\left(\{0\}\right) + \mu_{f_g(\alpha,U)}\left(\{1\}\right)\right) \\
&= \frac{1}{2}\alpha\left(g\left(\alpha\right) + 1 - g\left(2\alpha\right)\right) + \alpha\left(g\left(2\alpha\right) + 1 - g\left(\alpha\right)\right) \\
&= \frac{3}{2}\alpha + \frac{\alpha}{2}\left(g\left(2\alpha\right) - g\left(\alpha\right)\right).
\end{aligned}$$

To ensure $\mathbb{E}\left(G_{f_g}\right) > \frac{3}{2}\alpha$, it suffices to choose a measurable function g satisfying $g\left(2\alpha\right) > g\left(\alpha\right)$. Since the player does not know the exact value of α but knows that $\alpha < 2\alpha$, they can select any increasing (and hence measurable) function $g : \mathbb{R}_{>0} \to [0,1]$. Examples of such functions include:

$$g\left(x\right) = \frac{x}{x+1} = 1 - \frac{1}{x+1}, \quad g\left(x\right) = 1 - e^{-x}.$$

With these choices, we obtain $g\left(2\alpha\right) > g\left(\alpha\right)$, effectively improving the expected gain over the naive strategy by:

$$\frac{\alpha}{2}\left(g\left(2\alpha\right) - g\left(\alpha\right)\right).$$

For the above examples of g, this improvement is given by:

$$\frac{\alpha}{2}\left(\frac{1}{1+\alpha} - \frac{1}{1+2\alpha}\right), \quad \frac{\alpha}{2}\left(e^{-\alpha} - e^{-2\alpha}\right) \text{ respectively.}$$

Ideally, one would aim to find an increasing function

$$g \in \mathcal{G} := \{g' \mid g' : \mathbb{R}_{>0} \to [0,1] \text{ increasing}\}$$

that maximizes $\inf \{g(2x) - g(x) \mid x \in \mathbb{R}_{>0}\}$ however:

$$\sup \{\inf \{g'(2x) - g'(x) \mid x \in \mathbb{R}_{>0}\} \mid g' \in \mathcal{G}\} = 0$$

because if we assume for contradiction that there exists $c > 0$ and $g' \in \mathcal{G}$ such that

$$g'(2x) - g'(x) \geq c \quad \text{for all } x > 0.$$

Then by simply iterating this inequality we obtain by induction on $\mathbb{N}$:

$$g'(2^n x) - g'(x) \geq nc \quad \text{for all } n \in \mathbb{N}.$$

Since g' is bounded above by 1 and bounded below by 0, this implies $nc < 1$ for every n, which is impossible since c is not an infinitesimal element w.r.t to 1. However if we had more information on α (say some lower or/and upper bound) we may tune hyperparameter $t, c \in \mathbb{R}_{>0}$ by doing first order optimization on for examples:
Power-Law Functions:

$$g(x) = \frac{x^t}{x^t + c}.$$

For a given t, larger c makes it easier to switch for bigger and bigger x emphasizing to keep on only the very big x, smaller c makes it harder to switch for less and less big x emphasizing to keep most of the x. A fast transition occurs to the constant 1 when $c \to 0$. For a given c, larger t makes it harder to switch for less and less big x emphasizing to keep most of the x, smaller t makes it easier to switch for less and less big x emphasizing to keep only the very big x. A fast transitions to the constant $\frac{1}{c}$ occurs when $t \to 0$.
Logarithmic Functions:

$$g(x) = \frac{\ln(1 + x^t)}{\ln(1 + x^t) + c}.$$

This is exactly the same as above but things are more evenly spread out and smoother.
Exponential Functions:

$$g(x) = 1 - e^{-tx}.$$

Larger t makes it very difficult to switch for all x emphasizing to change only on the very small x, smaller t makes it easier to switch emphasizing only the very big x.

Remark. We can generalize the contents of the envelopes to be α and $k\alpha$ for any real number $k > 1$ and the previous analysis applies *mutatis mutandis*, where:

- The naive strategy yields an expected gain of $\frac{k+1}{2}\alpha$.
- The improved strategy, using any increasing function $g \colon \mathbb{R}_{>0} \to [0, 1]$, gives an expected gain of:

$$\frac{k+1}{2}\alpha + \frac{k-1}{2}\alpha\big(g(k\alpha) - g(\alpha)\big).$$

This setting is equivalent to considering α and $k\alpha$ where $0 < k \neq 1$ since we can simply swap the roles of the envelopes:

- If $k > 1$, the situation remains unchanged.
- If $0 < k < 1$, we reparameterize by setting $\alpha \leftarrow k\alpha$ and $k \leftarrow \frac{1}{k}$, reducing to the previous case.

In general, this framework is equivalent to considering any two distinct positive amounts α and β with $\alpha < \beta$, since we can express $\beta = k\alpha$ where $k = \frac{\beta}{\alpha} > 1$.

2 Problem A-3 (IMC 2018)

Determine all rational numbers a for which the matrix

$$A = \begin{bmatrix} a & a & 1 & 0 \\ -a & -a & 0 & 1 \\ -1 & 0 & a & a \\ 0 & -1 & -a & -a \end{bmatrix}$$

is the square of a matrix with all rational entries.

Answer:

We will show that the only such number is $a = 0$.

Let A be as given above, and suppose that $A = B^2$ for some matrix B with rational entries. It is easy to compute the characteristic polynomial of A, which is

$$p_A(X) = \det(A - XI) = \left(X^2 + 1\right)^2.$$

By the Cayley-Hamilton theorem, we have $p_A\left(B^2\right) = p_A(A) = 0$. Since $p_A\left(X^2\right) \in \mathbb{Q}[X]$ annihilates B, there must be a non-zero minimal polynomial $\mu_B(X) \in \mathbb{Q}[X]$ of B. We may assume $\mu_B(X)$ is monic. The minimal polynomial is irreducible over $\mathbb{Q}[X]$ and divide all polynomials with rationnal coefficient that vanish at B; in particular, $\mu_B(X)$ must be a divisor of the polynomial $p_A\left(X^2\right) = \left(X^4+1\right)^2$. So $\mu_B(X)$ must be a divisor of the polynomial X^4+1. However X^4+1 is the 8-th cyclotomic polynomial since:

$$\Phi_8(X) = \Phi_{2^3}(X) = \Phi_2\left(X^{\left(2^{3-1}\right)}\right) = X^4 + 1,$$

and therefore is irreducible over $\mathbb{Q}[X]$. Hence $\mu_B(X) = X^4 + 1$. Therefore,

$$A^2 + I = \mu_B(B) = 0.$$

Since we have

$$A^2 + I = \begin{bmatrix} 0 & 0 & 2a & 2a \\ 0 & 0 & -2a & -2a \\ -2a & -2a & 0 & 0 \\ 2a & 2a & 0 & 0 \end{bmatrix},$$

the equation $A^2 + I = 0$ forces $a = 0$.

In case $a = 0$, we have

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \end{bmatrix}^2,$$

hence $a = 0$ satisfies the condition.

3 Problem A-4 (IMC 2005)

Find all polynomials of degree $n \geq 1$

$$P(X) = \sum_{i=0}^{n} a_i X^i = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0 \quad (a_n \neq 0)$$

satisfying the following two conditions:

1. $\{a_i \mid i \in [\![0, n]\!]\} = [\![0, n]\!]$ and
2. all roots of $P(X)$ are rational numbers.

Answer:

Let $n \geq 1$. If $n = 1$, then trivially the only such polynomial is X. Else, $n \geq 2$, and let $P(X)$ be a polynomial of degree n satisfying the conditions. Note that $P(X)$ does not have any positive root because $P(x) > 0$ for every $x > 0$. Thus, we can represent the roots as $-\alpha_i$ for $i = 0, 1, \ldots, n-1$, where $\{\alpha_i \mid i \in n\} \subset \mathbb{Q}_+$.

If $a_0 \neq 0$, then the roots are strictly positive $\{\alpha_i \mid i \in n\} \subset \mathbb{Q}_{>0}$, and condition 1 gives that there exists some $k \in \mathbb{N}$ with $1 \leq k \leq n-1$ such that $a_k = 0$. Using Vieta's formulas for this coefficient a_k, we obtain:

$$\sum_{\substack{S \subset n \\ |S| = n-k}} \left(\prod_{i \in S} \alpha_i \right) = \frac{a_k}{a_n} = 0,$$

which is impossible since the left-hand side is strictly positive (as $k < n$). Therefore, $a_0 = 0$, and one of the roots of $P(X)$, say (without loss of generality) α_{n-1}, must be zero.

Consider the polynomial

$$Q(X) = \sum_{i=1}^{n} a_i X^{i-1} = a_n X^{n-1} + a_{n-1} X^{n-2} + \cdots + a_1.$$

One has $P(X) = XQ(X)$, so it has the zeros $-\alpha_i$ for $0 \leq i \leq n-2$. Again, using Vieta's formulas for the coefficients $\frac{a_1}{a_n}$, $\frac{a_2}{a_n}$ and $\frac{a_{n-1}}{a_n}$ (recall $n \geq 2$), we get:

$$\prod_{0 \leq i < n-1} \alpha_i = \frac{a_1}{a_n}, \quad \sum_{0 \leq i < n-1} \left(\prod_{\substack{0 \leq j < n-1 \\ j \neq i}} \alpha_j \right) = \frac{a_2}{a_n}, \quad \sum_{0 \leq i < n-1} \alpha_i = \frac{a_{n-1}}{a_n}.$$

Dividing the second equation by the first, we obtain

$$\frac{\sum_{0 \leq i < n-1} \left(\prod_{\substack{0 \leq j < n-1 \\ j \neq i}} \alpha_j \right)}{\prod_{0 \leq i < n-1} \alpha_i} = \sum_{0 \leq i < n-1} \frac{1}{\alpha_i} = \frac{a_2}{a_1}.$$

Applying the AM-HM inequality (which is a corollary of the very useful Generalized Mean Inequality):

$$\frac{a_{n-1}}{(n-1)a_n} = \frac{\sum_{0 \leq i < n-1} \alpha_i}{n-1} \geq \frac{n-1}{\sum_{0 \leq i < n-1} \frac{1}{\alpha_i}} = \frac{(n-1)a_1}{a_2}.$$

Rearranging, we find

$$\frac{a_2 a_{n-1}}{a_1 a_n} \geq (n-1)^2,$$

and since the ratio is trivially strictly bounded above by $\frac{n^2}{2}$, we have:

$$\frac{n^2}{2} > \frac{a_2 a_{n-1}}{a_1 a_n} \geq (n-1)^2 \implies n^2 + 2 < 4n,$$

which holds only for integers $n \leq 3$. Thus, the only possible polynomials $P(X)$ satisfying the given conditions are X, and otherwise they have degree 2 or 3, a constant term 0, and negative roots.

These polynomials can then be explicitly found by brute force (finitely many possibilities) and they form exactly the set:

$$\left\{X, \quad X^2 + 2X, \quad 2X^2 + X, \quad X^3 + 3X^2 + 2X, \quad 2X^3 + 3X^2 + X\right\}.$$

4 Problem A-6 (IMC 2005)

Let $m, n \in \mathbb{Z}$. Given a group G, denote by $G(m)$ the subgroup generated by the m-th powers of elements of G:

$$G(m) := \langle \{g^m \mid g \in G\} \rangle \leq G.$$

If $G(m)$ and $G(n)$ are commutative, prove that $G(\gcd(m,n))$ is also commutative. Here, $\gcd(m,n)$ denotes the greatest common divisor of m and n.

Answer:

If $m = 0$ or $n = 0$, this is trivial. Suppose now $|m|, |n| \geq 1$.

Recall that if H is a group and $H' \subset H$ is a subset, then the subgroup of H generated by H' is the smallest subgroup of H containing H', that is, $\langle H' \rangle = \bigcap_{H' \subset F \leq H} F$. It is easy to see that:

$$\langle H' \rangle = \left\{ \prod_{i \in k} g_i^{\epsilon_i} \;\middle|\; \exists k \in \mathbb{N} \, \exists g \in (H')^k \, \exists \epsilon \in \{-1, 1\}^k \right\}, \tag{1}$$

where the product is taken in the order of the integer $k = \{i \in \mathbb{N} \mid i < k\}$ and the empty product is e_H.

Write $d = \gcd(m, n)$. Notice that:

$$G(d) = \langle G(m) \cup G(n) \rangle .$$

Indeed, this follows from the monotonicity of $\langle \_ \rangle$ and the fact that every subgroup is a fixed point of $\langle \_ \rangle$. If $z \in \left\{ g^d \mid g \in G \right\}$, then $z = g^d$ for some $g \in G$. By Bézout's lemma, there exist two integers $l, r \in \mathbb{Z}$ with $lm + rn = d$, and we have:

$$z = g^d = g^{lm+rn} = \left((g^m)^{\operatorname{sign}(l)} \right)^{|l|} \left((g^n)^{\operatorname{sign}(r)} \right)^{|r|} .$$

Because $g^m \in G(m)$ and $g^n \in G(n)$, we have $z \in \langle G(m) \cup G(n) \rangle$. Since z was arbitrary, we get $\left\{ g^d \mid g \in G \right\} \subset \langle G(m) \cup G(n) \rangle$, and thus:

$$G(d) = \left\langle \left\{ g^d \mid g \in G \right\} \right\rangle \subset \langle G(m) \cup G(n) \rangle .$$

Similarly, let $z \in \{g^m \mid g \in G\}$. Then $z = g^m$ for some $g \in G$, and thus:

$$z = \left(g^d \right)^{\frac{m}{d}} \in G(d),$$

since $g^d \in G(d)$. As z was arbitrary, we conclude $\{g^m \mid g \in G\} \subset G(d)$. Thus, $G(m) \subset G(d)$. In the exact same manner, $G(n) \subset G(d)$, and therefore:

$$\langle G(m) \cup G(n) \rangle \subset G(d).$$

We also have:

$$\langle G(m) \cup G(n) \rangle = \langle \{g^m \mid g \in G\} \cup \{h^n \mid h \in G\} \rangle \,[3].$$

[3] If H is a group, I a set, and $\{H_i \mid i \in I\}$ are subgroups of H each generated by $\{S_i \mid i \in I\} \subset \mathscr{P}(H)$ respectively ($\forall i \in I, H_i = \langle S_i \rangle$), then $\left\langle \bigcup_{i \in I} H_i \right\rangle = \left\langle \bigcup_{i \in I} S_i \right\rangle$. Indeed, we clearly have $\bigcup_{i \in I} S_i \subset \bigcup_{i \in I} H_i$ (since $S_i \subset \langle S_i \rangle$) and thus $\left\langle \bigcup_{i \in I} S_i \right\rangle \subset \left\langle \bigcup_{i \in I} H_i \right\rangle$. If we take an element $z \in \bigcup_{i \in I} H_i$, then there exists $j \in I$ with $z \in H_j = \langle S_j \rangle \subset \left\langle \bigcup_{i \in I} S_i \right\rangle$, and so we have $\bigcup_{i \in I} H_i \subset \left\langle \bigcup_{i \in I} S_i \right\rangle$, which means that we have the other inclusion $\left\langle \bigcup_{i \in I} H_i \right\rangle \subset \left\langle \bigcup_{i \in I} S_i \right\rangle$.

It is also clear regarding equation (1) that if $S \subset G$ is constituted of elements that commute with one another, then $\langle S \rangle$ is a commutative subgroup (this follows from the fact that if $a, b \in S$ commute, then any two elements in $\{a^{-1}, b^{-1}, a, b\}$ commute). The converse of this statement is trivial. Therefore, by the two equalities above, showing commutativity of $G(d)$ is equivalent to showing commutativity of any two elements in $\{g^m \mid g \in G\} \cup \{h^n \mid h \in G\}$. Because we know that any two elements in $\{g^m \mid g \in G\}$ or $\{h^n \mid h \in G\}$ commute (since $G(m)$ and $G(n)$ are commutative), we only need to show that any element in $\{g^m \mid g \in G\}$ commutes with any other element in $\{h^n \mid h \in G\}$. So, without further ado, let any two generators a^m and b^n ($a, b \in G$). Showing commutativity of these two elements is equivalent to showing neutrality of their commutator:

$$z := [a^m, b^n] = a^{-m} b^{-n} a^m b^n.$$

Then the relations

$$z = \left(a^{-m} b a^m\right)^{-n} b^n = a^{-m} \left(b^{-n} a b^n\right)^m$$

show that $z \in G(m) \cap G(n)$. But then z is in the center of $G(d)$. Indeed to show $z \in Z(G(d))$, we let $x \in G(d) = \langle \{g^m \mid g \in G\} \cup \{h^n \mid h \in G\} \rangle$, then there exists a natural number k and a sequence of length $2k$ of element of the group $\mathbf{g} \in G^{2k}$ with:

$$x = g_0^m g_1^n \cdots g_{2k-2}^m g_{2k-1}^n,$$

and as $z \in G(m) \cap G(n)$, z commutes with any element in $\{g^m \mid g \in G\} \cup \{h^n \mid h \in G\}$ so we have by induction $zx = xz$. Now, from the relation $a^m b^n = b^n a^m z$, it easily follows by induction that for any integer $l \geq 0$, we have:

$$a^{ml} b^{nl} = b^{nl} a^{ml} z^{l^2}.$$

Indeed, for $l = 0$, we have:

$$a^{ml} b^{nl} = a^0 b^0 = e_G = b^0 a^0 z^0 = b^{nl} a^{ml} z^{l^2},$$

and for $l = 1$, we have:

$$a^{ml} b^{nl} = a^m b^n = b^n a^m z = b^{nl} a^{ml} z^{l^2},$$

where we used the above relation. Suppose this holds for $l \geq 1$. We show this holds for $l + 1$. We have:

$$a^{m(l+1)} b^{n(l+1)} = a^m \left(a^{ml} b^{nl}\right) b^n = a^m b^{nl} a^{ml} z^{l^2} b^n = a^m b^{nl} a^{ml} b^n z^{l^2},$$

where we use in the second equality the induction hypothesis and in the third the fact that z is in the center and hence z^{l^2} as well. Now:

$$a^m b^{nl} = b^n a^m z b^{n(l-1)} = b^n \left(a^m b^{n(l-1)}\right) z \overset{l-1 \text{ times}}{=\cdots=} b^{nl} a^m z^l,$$

where we do an induction on l by iteratively using the known relation and the fact that z is in the center. Thus:

$$a^{m(l+1)} b^{n(l+1)} = a^m b^{nl} a^{ml} b^n z^{l^2} = b^{nl} a^{m(l+1)} b^n z^{l^2+l},$$

where we used the two last results and the fact that z^l is in the center. Similarly:

$$a^{m(l+1)} b^n = \left(a^{ml} b^n\right) a^m z \overset{l \text{ times}}{=\cdots=} b^n a^{m(l+1)} z^{l+1},$$

where we do an induction on $l + 1$ by iteratively using the known relation and the fact that z is in the center. Thus:

$$a^{m(l+1)} b^{n(l+1)} = b^{nl} a^{m(l+1)} b^n z^{l^2+l} = b^{n(l+1)} a^{m(l+1)} z^{l^2+2l+1} = b^{n(l+1)} a^{m(l+1)} z^{(l+1)^2},$$

where we used the two last results and $l^2+2l+1=(l+1)^2$. This concludes the induction and proves the statement.

In particular, for any integer $l \geq 0$, we have:

$$z^{l^2} = a^{-ml}b^{-nl}a^{ml}b^{nl} = [a^{ml}, b^{nl}].$$

Setting the two integers $k := \frac{m}{d}$ and $k' := \frac{n}{d}$, since $nk = mk'$, we obtain that $a^{mk} = (a^k)^m$ and $b^{nk} = (b^{k'})^m$ i.e. $a^{mk}, b^{nk} \in G(m)$, and thus they commute by hypothesis, which means:

$$z^{k^2} = [a^{mk}, b^{nk}] = e_G.$$

Similarly, $a^{mk'} = (a^k)^n$ and $b^{nk'} = (b^{k'})^n$ i.e. $a^{mk'}, b^{nk'} \in G(n)$, and thus they commute by hypothesis, which means:

$$z^{k'^2} = [a^{mk'}, b^{nk'}] = e_G.$$

So $z^{\left(\frac{m}{d}\right)^2} = e_G = z^{\left(\frac{n}{d}\right)^2}$. Clearly, $\gcd(k,k')=1$, and thus $\gcd(k^2,k'^2)=1$, so by Bézout's lemma, there exist two integers $s,t \in \mathbb{Z}$ with $sk^2+tk'^2=1$. Hence:

$$e_G = (e_G)^s(e_G)^t = \left(z^{k^2}\right)^s \left(z^{k'^2}\right)^t = z^{sk^2+tk'^2} = z^1 = [a^m, b^n].$$

Since $a, b \in G$ were arbitrary, we conclude that $G(d)$ is commutative, as required.