Prerequisites for Cryptography & Security Course

Extended version

1 Probabilities

In this course we will work constantly with probabilities. Here we present a useful list of things you should know and be able to use.

Combinatorics.

We mostly use combinatorics in cryptography to look at the number of ways to arrange n different objects in a sequence.

We denote by $n!$ the factorial of n, i.e. $n! = 1 \cdot \ldots (n-1) \cdot n$. By convention we have $0! = 1$.

We usually use the following concepts

- permutations $P_n^k = \frac{n!}{(n-k)!}$ represent the number of ways to select k elements out of n where the order matters (i.e. choosing the values 1,2,3 is not the same as 1,3,2)
- combinations $C_n^k = \binom{n}{k} = \frac{n!}{(n-k)!k!}$ represent number of ways to select k elements out of n where the order does not matter
- combinations with repetitions $\binom{n+k-1}{k}$ is the same as combinations but now the selections are done with replacement

Events, random variable.

A random variable (r.v.) can take a set of different values, called support, with associated probability. In this course we work mostly with discrete random variables.The continuous random variables are usually measurements. An example of random variable X is the the outcome of rolling a die, i.e. $\Pr[X = i] = \frac{1}{6}$ for $1 \leq i \leq 6$. The support of X is the set $\{1, \ldots, 6\}$.

Independence of random variables.

Two random variables X and Y are independent when observing the outcome of X does not affect Y. In this case we have

$$\Pr[X = x, Y = y] = \Pr[X = x] \cdot \Pr[Y = y].$$

The random variables $X_1, \ldots, X_n$ are *pairwise independent* if any pair X_i and X_j, with $i \neq j$, is independent. They are *mutually independent* if for any subset $I \subseteq \{1, \ldots, n\}$ we have $\Pr[X_i = x_i, \text{ for all } i \in I] = \prod_{i \in I} \Pr[X_i = x_i]$.

Conditional probability and Bayes rule.

Conditional probability of r.v. X given another r.v. Y is defined by

$$\Pr[X = x | Y = y] = \frac{\Pr[X = x, Y = y]}{\Pr[Y = y]}.$$

When two r.v. X and Y are independent we have $\Pr[X = x | Y = y] = \Pr[X = x]$.

Bayes rule is defined as

$$\Pr[X = x | Y = y] = \frac{\Pr[Y = y | X = x] \cdot \Pr[X = x]}{\Pr[Y = y]}.$$

Law of total probability.

When trying to compute the probability of an event to happen it is sometimes handy to apply the law of total probability. Given the r.v. X and Y (whose support is the set S) we have that

$$\Pr[X = x] = \sum_{y \in S} \Pr[X = x | Y = y] \cdot \Pr[Y = y]$$

Inclusion-Exclusion principle.

For two r.v. X and Y we have that

$$\Pr[X = x \text{ or } Y = y] = \Pr[X = x] + \Pr[Y = y] - \Pr[X = x, Y = y].$$

Expected value.

Intuitively, expected value is the average value that a r.v. has for many repeated measurements. Given a r.v. whose support is a set S, the expected value is

$$E(X) = \sum_{x \in S} x \cdot \Pr[X = x].$$

For example, the expected value of X, where X is the outcome of a roll die, is $E(X) = \sum_{1 \leq i \leq 6} i \Pr[X = i] = 3.5$.

The expected value has the following properties:

- linearity: $E(c \cdot X + Y) = c \cdot E(X) + E(Y)$ and $E(c) = c$ for any r.v. X, Y and a constant c
- when two r.v. X and Y are independent: $E(X \cdot Y) = E(X) \cdot E(Y)$
- $E(f(X)) = \sum_{x \in S} f(x) \Pr[X = x]$

Variance.

Intuitively, the variance shows how spread out is the outcome from the expected value. Given a r.v. X, we have that

$$Var(X) = E((X - E(X))^2) = E(X^2) - E(X)^2.$$

The variance has the following properties:

- $Var(X) \geq 0$
- $Var(a \cdot X + b) = a^2 \, Var(X)$ for any constants a, b
- when two r.v. are independent: $Var(X + Y) = Var(X) + Var(Y)$
- when two r.v. are independent: $Var(X \cdot Y) = E(X^2)E(Y^2) - E(X)^2 E(Y)^2$

Distributions.

We give a list of useful distributions:

- **Bernoulli distribution**: takes a single parameter p and is denoted $\text{Ber}(p)$. It can be seen as the flipping of a biased coin where $\Pr[X = 1] = p$ and $\Pr[X = 0] = 1 - p$, with $0 \leq p \leq 1$. We have $E(X) = p$ and $Var(X) = p(1 - p)$.
- **Binomial distribution**: takes two parameters p and n and is denoted $\text{Bi}(n, p)$. It represents the sum of n independent biased coins where $\Pr[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}$.
  We have $E(X) = np$ and $Var(X) = np(1 - p)$.
- **Geometric distribution**: takes a single parameter p and is denoted $\text{Ge}(p)$. It is defined as the "number of trials until you succeed" to get a 1 from a biased coin, i.e. $\Pr[X = k] = (1 - p)^{k-1} p$.
  We have $E(X) = \frac{1}{p}$ and $Var(X) = \frac{1-p}{p^2}$.

Indicator function.

For a r.v. X and a predicate P we define the indicator function as

$$1_{P(X)} \coloneqq \begin{cases} 1 & \text{if P(X)=}\mathsf{true} \\ 0 & \text{otherwise} \end{cases}$$

We have that $E(1_{P(X)}) = \Pr[P(X) = \mathsf{true}]$. A simple example for P is $P : X = a$ for some a in support of X.

Inequalities.

Some inequalities that can be useful are the Jensen, Chebyshev and Markov inequalities.

2 Proof Techniques

Proofs are one of the important parts of this course. It is very important to be formal and clear while you are proving a statement. Basically, you should know the basic proof techniques below:

Proof by Contradiction. Given a statement P, if we want to prove that P is true, we can first assume that P is false. We then try to deduce a contradiction, i.e. find a series of implications $\neg P \Rightarrow \ldots \Rightarrow \neg Q$ for some statement Q which we know is true. If we succeed, then clearly $\neg P$ is wrong, and so P must be true.

Example: $\forall n \in \mathbb{Z}$, if n^2 is odd, then n is odd.

Suppose it is not the case which means we assume that the negation of the statement (n^2 is odd and n is even) is true.

Since n is even, $\exists k \in \mathbb{Z}$ such that $n = 2k$. So, we can write $n^2 = (2k)(2k) = 4k^2 = 2(2k^2)$. Because $2, k \in \mathbb{Z}$, the products of them are in $\mathbb{Z}$ ($2k^2 \in \mathbb{Z}$). From the definition of even number, n^2 *is even which contradicts with our assumption.* It means that the negation of the statement is false so the original statement is true.

Proof by contrapositive. Given a statement $P \Rightarrow Q$, we can show the contrapositive of the statement which is $\neg Q \Rightarrow \neg P$ is true to show the statement is true.[1]

Example: Let $a, b \in \mathbb{Z}$. If $n \nmid ab$, then $n \nmid a$ and $n \nmid b$.

Contrapositive: If $n \mid a$ or $n \mid b$, then $n \mid ab$.

Suppose $n \mid a$. Then, $\exists k \in \mathbb{Z}$ such that $a = nk$. We can write $ab = nkb = n(kb)$. So, $n \mid ab$. Similarly, suppose $n \mid b$. Then, $\exists \ell \in \mathbb{Z}$ such than $b = n\ell$. We can write $ab = an\ell = n(\ell a)$. So, $n \mid ab$. Lastly, suppose $n \mid a$ and $n \mid b$. Then, $\exists k, \ell \in \mathbb{Z}$ such than $a = nk$ and $b = n\ell$. We can write $ab = nkn\ell = n(k\ell n)$. So, $n \mid ab$.

Proof by induction. This proof technique is used to show the statements such as $P(n)$ is correct for any $n \geq c$ where $c \in \mathbb{N}$. You should always use the following template in induction proofs:

1. **Base case:** One or more particular cases that represent the most basic case, e.g. showing $P(n)$ is true for $n = 1$.
2. **Induction hypothesis:** Assumption that we want to be based on, e.g. assume that $P(k)$ is true.
3. **Inductive step:** Proving the next step based on the induction hypothesis, e.g. prove $P(k)$ implies $P(k + 1)$.

[1] If you're not sure that an implication and its contrapositive are equivalent, you can prove it as an exercise.

Example: Let $n \in \mathbb{N}$. Prove that $\sum_{i=0}^{n} i = \frac{n(n+1)}{2}$, i.e. $P(n) : \sum_{i=0}^{n} i = \frac{n(n+1)}{2}$.

1. **Base case:** For $n = 0$, $\sum_{i=0}^{0} i = \frac{0(0+1)}{2} = 0$. So, for $n = 0$, the statement is true ($P(0)$ is true).
2. **Induction hypothesis:** Assume that for $k \in \mathbb{N}$, $\sum_{i=0}^{k} i = \frac{k(k+1)}{2}$ ($P(k)$ is true).
3. **Inductive step:** We know that $P(k)$ is true where $k \in \mathbb{N}$. We want to prove that $P(k+1)$ is true.

$$\sum_{i=0}^{k+1} i = (k+1) + \sum_{i=0}^{k} i = (k+1) + \frac{k(k+1)}{2} = \frac{k^2+3k+2}{2} = \frac{(k+1)(k+2)}{2}$$

So, $P(k+1)$ is true.

General mistakes in proofs.

Do not forget the following remarks while you are proving a statement:

- You have to be clear in your statements. For example, "if n is even then $n = 2k$". **This statement is not complete.** You have to specify the domain of k i.e. $k \in \mathbb{Z}$.
- You **cannot give an example** to show that the statement is true. For example, **must definitely avoid** a proof attempts such as "$P(n)$ is true for $n = 1$, $n = 2$, $n = 8$, then $P(n)$ is true for all n".
  However, you can show a statement is false by giving a counterexample. For example, consider the following statement: 'all prime numbers are odd'. You can give as a counterexample '2' which is even and prime number **to disprove the statement.**
- You must have **clear explanations** in the proofs. For example, if you say P implies Q, you should explain what is the reason of this implication. Do not just say "it is trivial that..." If something is trivial, tell us why in a sentence. Besides, if you use a known theorem, definition or lemma, you must give its name, e.g. because of Lagrange theorem, $P(n)$ implies $Q(n)$.
- If you are required to prove an equivalence, i.e. a statement of the form "P if and only if Q", then you have to show **2 implications:** $P \Rightarrow Q$ **and** $Q \Rightarrow P$.

3 Algorithms

Complexity of an Algorithm.

Definition 1 ($O(.)$). *Assume that $f(n)$ and $g(n)$ are two functions. We write $f(n) = O(g(n))$ if and only if there exist constants $N > 0$ and $C > 0$ such that $|f(n)| \leq C|g(n)|$ for all $n \geq N$.*

Informally, g defines an upper bound for f. In other words, the growth rate of f is less (or equal) than the growth rate of g.

Example: $f(n) = 4n^3 + 2n^2 + 5n + 6$ is $O(n^3)$ with the constants $C = 17$ and $N = 1$.

Definition 2 ($\Omega(.)$). *Assume that $f(n)$ and $g(n)$ are two functions. We write $f(n) = \Omega(g(n))$ if and only if there exist constants $N > 0$ and $C > 0$ such that $|f(n)| \geq C|g(n)|$ for all $n \geq N$.*

Informally, g defines a lower bound for f. In other words, the growth rate of f is greater (or equal) than the growth rate of g.

Example: $f(n) = 5n^2$ is $\Omega(n)$ with $C = 5$ and $N = 1$.

Definition 3 ($\Theta(.)$). *Assume that $f(n)$ and $g(n)$ are two functions. We write $f(n) = \Theta(g(n))$ if and only if there exist constants $C_1, C_2 > 0$ and $N > 0$ such that $C_1|g(n)| \leq |f(n)| \leq C_2|g(n)|$.*

In other words, $g(n)$ is the exact bound for $f(n)$.

Example: $f(n) = n + 5n^{0.5}$ is $\Theta(n)$ with the constants $C_1 = 1$, $C_2 = 6$ and $N = 1$.

As an exercise, show that $f(n) = \Theta(g(n))$ is equivalent with $f(n) = O(g(n)) \wedge f(n) = \Omega(g(n))$.

Definition 4 ($o(.)$). *Assume that $f(n)$ and $g(n)$ are two functions. We write $f(n) = o(g(n))$ if and only if*

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0.$$

Randomized Algorithms. Randomized algorithms flip a coin to determine what is the next step in the execution. We can analyze *expected complexity* in them since execution of time may differ according to flipped coins. In addition, the outcome of the randomized algorithm may also differ in each execution so we can only compute *expected outcome*.

Consider the following example:

$x = 0$
for $i = 1$ **to** 10 :
$b \leftarrow$ flip coin
  if $b = head$
    $x = x + 2$
  else:
    $x = x - 3$

Expected number of addition (subtraction is the same):

$$E(\#addition) = \sum_{i=1}^{10} 1 \cdot \Pr[b = head] = 5,$$

because there is one addition every time $b = head$.

Expected Output:

$$E(output) = \sum_{i=1}^{10} 2 \cdot \Pr[b = head] + (-3) \cdot \Pr[b = tail] = -5,$$

because we add a 2 whenever $b = head$ and we subtract 3 otherwise.

We may consider the worst-case performance of an algorithm. As an example, we can give the similar example as above with a "for loop" in one of the if statement i.e "if $b = head$". In the worst case, we always have head. Then, the complexity of the algorithm will be quadratic while in the best case, we will have always tail and the complexity of the algorithm will be linear.

4 Algebra

Algebraic structures are of great importance in cryptography. All relevant public key cryptosystems rely on properties of certain algebraic structures, and algebraic structures and their properties can also be used to attack/prove security of other cryptographic primitives. We list here the most important concepts from algebra, that you should already be familiar with.

Gcd.

Definition 5. *For two integers a, b where w.l.o.g. $b \neq 0$, we call the integer d the* greatest common divisor of a and b *if and only if for any integer c that divides both a and b, c also divides d. We denote it $d = \gcd(a, b)$.*

This is equivalent to saying, that $\gcd(a,b) = \max\{i \in \mathbb{Z} \mid i \textit{ divides } a \text{ and } i \textit{ divides } b\}$. Here are several properties of the gcd of non-zero integers a, b, c, which are useful to know:

- $\gcd(a,0) = a$, $\gcd(a,1) = 1$
- $\gcd(ca,cb) = c \cdot \gcd(a,b)$
- $a \mid b \Leftrightarrow \gcd(a,b) = a$
- $\gcd(a,b) \mid \gcd(a,bc)$
- $\gcd(a,b) = \gcd(a, b+ca)$

Congruences.

Definition 6. *Given a positive integer n, we say that two integers a, b are* congruent modulo *n, if n divides $(a-b)$ and denote this as $a \equiv b \pmod{n}$.*

Note, that if $a \equiv b \pmod{n}$, then we have $a = b + kn$ for some $k \in \mathbb{Z}$.

We call an equation of the form

$$a \cdot x \equiv b \pmod{n}$$

with $a, b, n \in \mathbb{Z}$ and n positive a *linear congruence*. A linear congruence does not always have a solution:

- If $\gcd(a,n) = 1$, then there is exactly one solution in the set $\{0, 1, \ldots, n-1\}$: $x = b \cdot a^{-1} \bmod n$.
- If $d = \gcd(a,n) > 1$ and $d \nmid b$, then there is no solution.
- If $d = \gcd(a,n) > 1$ and $d \mid b$, then there are d solutions in the set $\{0, 1, \ldots, n-1\}$: We solve the congruence $\frac{a}{d} \cdot x' \equiv \frac{b}{d} \pmod{\frac{n}{d}}$ as in the first case, and find solutions modulo n as $x = x' + k \cdot \frac{n}{d}$ for $k \in \{0, 1, \ldots, d-1\}$.

Here, $a^{-1} \bmod n$ is an integer called *multiplicative inverse of a modulo n*.

It is defined as the integer, for which $a \cdot a^{-1} \equiv 1 \pmod{n}$, i.e. $a \cdot a^{-1} = 1 + k \cdot n$ for some integer k. Note that if $\gcd(a,n) > 1$, then $a^{-1} \bmod n$ does not exist! To compute modular inverses, we generally use the Extended Euclidean algorithm (EEA). However, for computing the modular inverses when solving linear congruences with small n ($n < 200$) by hand, the EEA is not useful. Here, it is faster to try all possibilities (for very small n only!) or try to be clever (e.g. try dividing increasing multiples of n augmented by 1 by a).

Extended Euclidean algorithm. The extended Euclidean algorithm is an efficient algorithm, that takes two integers a, b on input and computes both $\gcd(a,b)$, and integers u, v, such that $a \cdot u + b \cdot v = \gcd(a,b)$ (This property is called Bézout's identity, and can be used to compute the modular inverses!). We omit the pseudocode of the algorithm in this sheet, however it is one of the very few algorithms **you must know by heart**. Instead of memorizing it as pseudocode, it might be easier to grasp the intuition through few exercises[2].

[2] http://www-math.ucdenver.edu/~wcherowi/courses/m5410/exeucalg.html

Groups.

Definition 7. *A group is a set G together with a mapping from $G \times G$ to G which maps (a, b) to an element denoted $a \odot b$ and such that*

1. *[closure] for any $a, b \in G$, we have $a \odot b \in G$,*
2. *[associativity] for any $a, b, c \in G$, we have $(a \odot b) \odot c = a \odot (b \odot c)$*
3. *[neutral element] there exists an element $e \in G$ s.t. for any a, $a \odot e = e \odot a = a$,*
4. *[invertibility] for any $a \in G$ there exists $b \in G$ s.t. $a \odot b = b \odot a = e$.*

Examples of groups are integers with addition $(\mathbb{Z}, +)$, non-zero rationals with multiplication $(\mathbb{Q}\backslash\{0\}, \times)$, or the set of all permutations of integers $\{0, \ldots, n-1\}$ for some n, together with composition of functions $(S_n, \circ)$.

Definition 8. *An Abelian group is a set G together with a mapping from $G \times G$ to G which maps (a, b) to an element denoted $a \odot b$ and such that*

1-4. *[group] it is a group,*
5. *[commutativity] for any a, b we have $a \odot b = b \odot a$.*

The groups $(\mathbb{Z}, +)$ and $(\mathbb{Q}\backslash\{0\}, \times)$ are Abelian, while $(S_n, \circ)$ is not.
You should also be familiar with the notion of a *subgroup*: given a group G, we call subset $H \subseteq G$ a subgroup of G, if H itself is a group with respect to the same operation as used in G (i.e. it verifies all 4 properties of a group). Note that every group has at least two trivial subgroups: the whole group itself, and a singleton set containing the neutral element.

The group operation is very frequently denoted either with *additive* notation (then we write $a + b$ for the operation, $-a$ for the inverse elements, and 0 for the neutral element) or with *multiplicative* notation (then we write ab for the operation, a^{-1} for the inverse elements, and 1 or e for the neutral element).

Order of a group, order of an element. Given a finite group G (with multiplicative notation), we call its cardinality $|G|$ the *order of* G. Given an element $a \in G$, we call the smallest positive integer d, s.t. $a^d = 1$ the *order of* a, i.e. $\text{ord}(a) = \min\{d \,|\, a^d = 1 \wedge d \geq 1\}$. The order of an element a is in fact the order of a subgroup $\langle a \rangle \subseteq G$ generated by a.

Lagrange Theorem.

Theorem 1. *Let G be a finite group. Then, for every $g \in G$, we have that ord(g) divides order of G.*

We omit the proof here, we mostly use this theorem for deriving further properties (although it is good to know how the proof works as well). As a direct consequence, we have that for all $g \in G$, we have $g^{|G|} = 1$. This in particular implies $g^n = g^{n \bmod |G|}$ for all $g \in G$ and $n \in \mathbb{Z}$.

Cyclic groups and generators. If there exists an element g in a finite group G, such that for all $h \in G$ we can write $h = g^i$ for some $i \in \mathbb{Z}$, then we call G a *cyclic group* and g a *generator of* G. We then have $\text{ord}(g) = |G|$. Note that every cyclic group is necessarily Abelian. An example of a cyclic group is $\mathbb{Z}_n$, i.e. the set of non-negative integers smaller than n together with addition modulo n (a generator is for example 1).

Rings and Fields.

Definition 9. *A ring is an Abelian group $(R, +)$ together with a mapping from $R \times R$ to R which maps (a, b) to an element denoted ab and such that*

1-4. [group] R with $+$ is a group,
5. [Abelian] for any $a, b \in R$, we have $a + b = b + a$,
6. [closure] for any $a, b \in R$, we have $ab \in R$,
7. [associativity] for any a, b, c, we have $(ab)c = a(bc)$,
8. [neutral element] there exists $1 \in R$ s.t. for any $a \in R$, $a1 = 1a = a$,
9. [distributivity] for any $a, b, c \in R$, we have $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$.

Definition 10. *A commutative ring is a ring R such that*

1-9. [ring] it is a ring,
10. [commutativity] for any $a, b \in R$ we have $ab = ba$.

Examples of commutative rings are integers with addition and multiplication $(\mathbb{Z}, +, \times)$, or univariate polynomials of finite degree with real coefficients with addition and multiplication of polynomials $(\mathbb{R}[x], +, \times)$.
You should also be familiar with the notion of an *ideal* (we limit ourselves to commutative rings for brevity): given a commutative ring R, we call a subset $I \subseteq R$ an ideal of R, if I itself is a ring with respect to the the original operations $+, \cdot$ of R, and if additionally $ai \in I$ for every $a \in R$ and $i \in I$.

Definition 11. *A field is a commutative ring $(K, +, \times)$ such that*

1-9. [ring] K is a ring with $+$ and $\times$,
10. [commutativity] for any $a, b \in K$, we have $ab = ba$,
11. [invertibility] for any $a \in K \backslash \{0\}$ there exists $b = a^{-1}$ s.t. $ab = ba = 1$.

Here 0 denotes the neutral element of the operation $+$. An example of a field are the reals with addition and multiplication $(\mathbb{R}, +, \times)$.

Homomorphisms and isomorphisms.

Definition 12. *Given two groups $(G, \odot), (H, \boxdot)$, we call a function $f : G \to H$ a* group homomorphism*, if for all $g, g' \in G$ we have*

$$f(g \odot g') = f(g) \boxdot f(g').$$

If, in addition, f is also a bijection, we call f an group isomorphism*.*

This means, that a group homomorphism is a function which preserves some structure shared between the two group operations. An example of a group homomorphism is the sign function, that assigns a value $\operatorname{sgn}(q) \in \{-1, 1\}$ to a non-zero rational number q, depending on its sign (i.e. $\operatorname{sgn}(q) = \frac{q}{|q|}$). The function $\operatorname{sgn}(\cdot)$ is in fact a homomorphism between $(\mathbb{Q}\backslash\{0\}, \times)$ and $(\{1, -1\}, \times)$.

An example of an isomorphism is the natural logarithm $\log : \mathbb{R}^+ \to \mathbb{R}$. Indeed, every positive real number has a unique logarithm, and we have $\log(ab) = \log(a) + \log(b)$.

Euler theorem, little Fermat theorem.

Theorem 2 (little Fermat). *For a prime p and an $a \in \mathbb{Z}$ s.t. $\gcd(a, p) = 1$ we have*

$$a^{p-1} \equiv 1 \pmod{p}.$$

The Euler theorem is a generalization of the little Fermat's theorem.

Theorem 3 (Euler). *For a positive integer n and an $a \in \mathbb{Z}$, s.t.* $\gcd(a, n) = 1$, *we have*

$$a^{\varphi(n)} \equiv 1 \pmod{n}.$$

Here, the function $\varphi(n)$ returns the number of positive integers smaller than n, that are also coprime with n. It is computed as

$$\varphi(n) = \begin{cases} p^{\alpha-1}(p-1) & \text{for } n = p^\alpha \text{ with } p \text{ prime and } \alpha \geq 1 \\ \prod_{i=1}^{r} \varphi(p_i^{\alpha_i}) & \text{for } n = \prod_{i=1}^{r} p_i^{\alpha_i} \text{ with } p_1, \ldots, p_r \text{ pairwise different primes.} \end{cases}$$

Among other things, these two theorems can be used to efficiently compute powers of integers modulo n. For example

$$2^{34} \bmod 31 = 2^{34 \bmod 30} \bmod 31 = 2^4 \bmod 31 = 16$$

because $\varphi(31) = 30$ and 2 and 31 are coprime. However

$$2^{33} \bmod 4 \neq 2^{33 \bmod \varphi(4)} \bmod 4$$

because $\gcd(2, 4) > 1$! We have

$$2^{33} \bmod 4 = 0$$

while

$$2^{33 \bmod \varphi(4)} \bmod 4 = 2^{33 \bmod 2} \bmod 4 = 2.$$