Probability Distribution: Building up the notion of Pseudo-randomness

Debdeep Mukhopadhyay
IIT Kharagpur

Probability Distribution

1. Probability Distribution: $p = (p_1, ..., p_n)$ is a tuple of elements $p_i \in R_n$, $0 \le p_i \le 1$, called probabilities, such that $\sum_{i=1}^{n} p_i = 1$.

2. A probability space (X, p_X) is a finite set $X = \{x_1, ..., x_n\}$ equipped with a probability distribution $p_X = \{p_1, ..., p_n\}$.

p_i is called the probability of x_i, $1 \le i \le n$. We also write $p_X(x_i) = p_i$ and consider p_X as a map $X \to [0,1]$, called the probability measure on X, associating with $x \in X$ its probability.

3. An event ε in a probability space (X, p_X) is a subset ε of X.

$$p_X(\varepsilon) = \sum_{y \in \varepsilon} p_X(y)$$

$$\therefore p_X(X) = 1$$

A probability space X is the model of a random experiment.
n independent repetitions of the random experiment are modeled by the direct product: $X^n = X \times X \times ... \times X$

Some interesting results…

Let ε be an event in a probability space X, with $\Pr[\varepsilon]=p>0$. Repeatedly, we perform the random experiment X independently. Let, G be the expected number of experiments of X, until ε occurs the first time. Prove that: $E(G)=\frac{1}{p}$

$$\Pr[G=t]=(1-p)^{t-1}p \Rightarrow E(G)=\sum_{t=1}^{\infty} tp(1-p)^{t-1} = -p\frac{d}{dp}\sum_{t=1}^{\infty}(1-p)^t = -p\frac{d}{dp}(\frac{1}{p}-1) = \frac{1}{p}.$$

Another Useful result

Let R, S and B be jointly distributed r.v with values in {0,1}.
Assume that B and S are independent and that B is uniformly distributed:
Pr(B=0)=Pr(B=1)=1/2
Prove that: Pr(R=S)=1/2 + Pr(R=B|S=B)-Pr(R=B)

$$\Pr(S{=}B)=\Pr(S{=}0)\Pr(B{=}0|S{=}0)+\Pr(S{=}1)\Pr(B{=}1|S{=}1)$$
$$=\Pr(S{=}0)\Pr(B{=}0)+\Pr(S{=}1)\Pr(B{=}1)$$
$$=\frac{1}{2}(\Pr(S{=}0)+\Pr(S{=}1))=\frac{1}{2}$$
$$\textit{Likewise}, \Pr(S=\overline{B})=\frac{1}{2}$$
$$\Pr(R=S)=\frac{1}{2}\Pr(R=B\mid S=B)+\frac{1}{2}\Pr(R=\overline{B}\mid S=\overline{B})$$
$$=\frac{1}{2}[\Pr(R=B\mid S=B)+1-\frac{1}{2}\Pr(R=B\mid S=\overline{B})]$$
$$=\frac{1}{2}+\frac{1}{2}[\Pr(R=B\mid S=B)-\frac{\Pr[(R{=}B)\cap(S{=}\overline{B})]}{\Pr(S=\overline{B})}]$$
$$\because (R{=}B)=((R{=}B)\cap(S{=}\overline{B}))\cup((R=B)\cap(S=B))$$
$$\therefore \Pr[R=B]=\Pr[(R{=}B)\cap(S{=}\overline{B})]+\Pr[(R=B)\cap(S=B)]$$
$$\Rightarrow\quad \Pr(R=S)=\frac{1}{2}+\frac{1}{2}(\Pr(R=B\mid S=B)-\frac{\Pr[R=B]-\Pr[(R=B)\cap(S=B)]}{\Pr(S=\overline{B})})$$
$$=\frac{1}{2}+\frac{1}{2}(\Pr(R=B\mid S=B)-\frac{\Pr[R=B]-\Pr[S=B]\Pr[(R=B)\mid(S=B)]}{1/2})$$
$$=\frac{1}{2}+\frac{1}{2}(\Pr(R=B\mid S=B)-\frac{\Pr[R=B]-1/2\Pr[(R=B)\mid(S=B)]}{1/2})$$
$$=\frac{1}{2}+\Pr(R=B\mid S=B)-\Pr[R=B]$$

Statistical Distance between Probability Distributions

Let p and $\tilde{p}$ be probability distributions on a finite set X.

The statistical distance between p and $\tilde{p}$ is:

$$\text{dist}(p,\tilde{p}) = \frac{1}{2}\sum_{x\in X} | p(x) - \tilde{p}(x) |$$

The statistical distance between probability distributions p and $\tilde{p}$ on a finite set X is the maximal distance between the probabilities of events in X, ie.

$$\text{dist}(p,\tilde{p}) = \max_{\varepsilon\subseteq X} | p(\varepsilon) - \tilde{p}(\varepsilon)|$$

The events in X are the subsets of X. We divide the subsets into three categories:

$$\varepsilon_1 = \{x \in X \mid p(x) > \tilde{p}(x)\}$$

$$\varepsilon_2 = \{x \in X \mid p(x) < \tilde{p}(x)\}$$

$$\varepsilon_3 = \{x \in X \mid p(x) = \tilde{p}(x)\}$$

We have $0 = p(X) - \tilde{p}(X) = \sum_{i=1}^{3}[p(\varepsilon_i) - \tilde{p}(\varepsilon_i)]$

$\because p(\varepsilon_3) - \tilde{p}(\varepsilon_3) = 0 \Rightarrow p(\varepsilon_1) - \tilde{p}(\varepsilon_1) = -(p(\varepsilon_2) - \tilde{p}(\varepsilon_2))$

Now because of the definition of ε_1,

$\max_{\varepsilon\subseteq X}|p(\varepsilon)-\tilde{p}(\varepsilon)| = p(\varepsilon_1) - \tilde{p}(\varepsilon_1) = -(p(\varepsilon_2) - \tilde{p}(\varepsilon_2))$

$$\therefore \text{dist}(p,\tilde{p}) = \frac{1}{2}\sum_{x\in X} | p(x) - \tilde{p}(x) |$$

$$= \frac{1}{2}\left(\sum_{x\in\varepsilon_1}[p(x) - \tilde{p}(x)] - \sum_{x\in\varepsilon_2}[p(x) - \tilde{p}(x)]\right)$$

$$= \frac{1}{2}[(p(\varepsilon_1) - \tilde{p}(\varepsilon_1)) - (p(\varepsilon_2) - \tilde{p}(\varepsilon_2))] = \max_{\varepsilon\subseteq X} | p(\varepsilon) - \tilde{p}(\varepsilon)|$$

Indistinguishable Distributions

p and $\tilde{p}$ are called polynomially close or ε-indistinguishable if:

$$\text{dist}(p,\tilde{p}) \le \varepsilon(n) = \frac{1}{P(n)}$$

where $\varepsilon(n)$ is a negligible quantity. p(n) is a polynomial in n.

Pseudo-random sequence: No efficient observer can distinguish it from a uniformly chosen string of the same length.

This approach leads to the concept of pseudo-random generators, which is a fundamental concept with lot of applications.

Proof

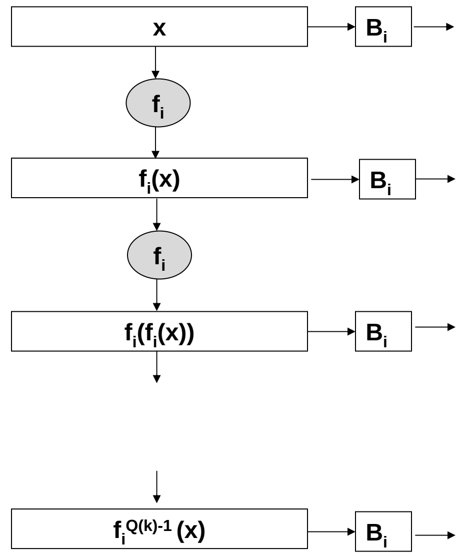
Let $J_k = \{n \mid n = rs, r, s \text{ are primes}, |r|=|s|=k, r \neq s\}$ and $x \leftarrow Z_n$ and $x \leftarrow Z_n^*$ are polynomially close. Is the result dependent on the choice of r and s?

Pseudorandom Bit Generator

- Let $I=(I_n)_{n \in N}$ be a key set with security parameter n, and let K be a probabilistic sampling algorithm for I, which on input (n) outputs an $i \in I_n$. Let l be a polynomial function in the security parameter.
- A pseudorandom bit generator with key generator K and stretch function *l* is a family of functions $G=(G_i)_{i \in I}$ of functions.
  - $G_i: X_i \rightarrow \{0,1\}^{l(n)}$, $i \in I(n)$
  - G is computable by a deterministic polynomial algorithm G.
    - $G(i,x)=G_i(x)$ for all $i \in I$ and $x \in X_i$
    - there is a uniform sampling algorithm for X. On input i, it outputs $x \in X_i$.

Pseudorandom Bit Generator

$$
\begin{aligned}
&|\Pr(A(i,z)=1 : i = K(1^n), z \leftarrow \{0,1\}^{l(n)} \\
&-\Pr\left(A(i,G_i(x)\right)=1) : i = K(1^n), x \leftarrow X_i \,| \\
&\leq \frac{1}{P(n)}
\end{aligned}
$$



If the discrete log assumption is true,

$$Exp = (Exp_{p,g} : Z_{p-1} \to Z_p^*, x \to g^x \bmod p)$$

with I={(p,g)|p is prime, $g \in Z_p^*$ a primitive root}

is a bijective one-way function.

$$\text{MSB}_p(x) = \begin{cases} 0 \text{ for } 0 \le x < (p-1)/2 \\ 1 \text{ for } (p-1)/2 \le x \le p-1 \end{cases}$$

is a hard-core predicate for Exp.

Exp can be treated as a one-way permutation,

identifying Z_{p-1} with Z_p^*.

$$Z_{p-1} = \{0, ..., p-2\}$$

$$Z_p^* = \{1, ..., p-1\}$$

using the mapping $0 \to$ p-1, $1 \to 1$, ...,p-2 $\to$ p-2

Induced PRG is a called Blum Micali Generator.

Blum-Micali-Yao's Theorem

- Suppose f is a length preserving one-way function. Let B be a hard core predicate for f. Then the algorithm G defined by G(x)=F(x)||B(x)=F(x).B(x) is a pseudo random generator.

Let D be an algorithm distinguishing between $G(U_n)$ and U_{n+1}.

$\therefore \Pr[D(G(U_n)) = 1] - \Pr[D(U_{n+1}) = 1] > \varepsilon$

Define: $E^{(1)} = [f(U_n).b(U_n)]$

$E^{(2)} = [f(U_n).\overline{b}(U_n)]$

Note: $G(U_n) = f(U_n).b(U_n) = E^{(1)}$

$$
\begin{aligned}
&Also, \Pr[D(U_{n+1}) = 1] \\
&= \Pr[D(f(U_n).U_1) = 1][as,\ f \text{ is bijective}] \\
&= \Pr[D(f(U_n).b(U_n)) = 1]\Pr[b(U_n) = U_1] \\
&+ \Pr[D(f(U_n).\bar{b}(U_n)) = 1]\Pr[\bar{b}(U_n) = U_1] \\
&= \frac{1}{2}(\Pr[D(f(U_n).b(U_n)) = 1] + \Pr[D(f(U_n).\bar{b}(U_n)) = 1]) \\
&= \frac{1}{2}(\Pr[D(\mathrm{E}^{(1)}) = 1] + \Pr[D(\mathrm{E}^{(2)}) = 1])
\end{aligned}
$$

$$
\begin{aligned}
&\therefore \Pr[D(G(U_n)) = 1] - \Pr[D(U_{n+1}) = 1] \\
&= \Pr[D(E^{(1)} = 1] - \frac{1}{2}(\Pr[D(\mathrm{E}^{(1)}) = 1] + \Pr[D(\mathrm{E}^{(2)}) = 1]) \\
&= \frac{1}{2}(\Pr[D(E^{(1)} = 1] - \Pr[D(\mathrm{E}^{(2)}) = 1]) > \varepsilon
\end{aligned}
$$

Thus using D if we make an algorithm to guess the hardcore predicate B(.) from y=f(x), then we are done.

Algorithm A:

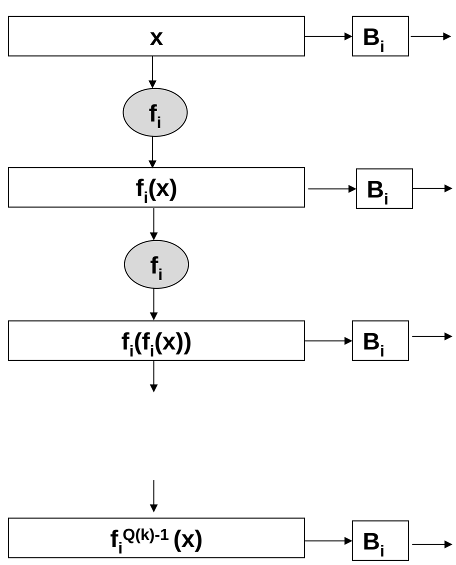
1. Select σ uniformly in {0,1}
2. If $D(y.\sigma)=1$, output σ, else $1-\sigma$

What is the probability that A is able to compute the hardcore predicate?:

$$\Pr[A(f(X)=b(X)]=\Pr[A(f(U_n)=b(U_n)]$$
$$=\Pr[D(f(U_n)U_1)=1 \wedge U_1=b(U_n)]$$
$$+\Pr[D(f(U_n)U_1)=0 \wedge 1-U_1=b(U_n)]$$
$$=\frac{1}{2}(\Pr[D(f(U_n)b(U_n))=1]$$
$$+\Pr[D(f(U_n)\bar{b}(U_n))=0])$$
$$=\frac{1}{2}(\Pr[D(f(U_n)b(U_n))=1]$$
$$+\frac{1}{2}(1-\Pr[D(f(U_n)\bar{b}(U_n)]=1)$$
$$=\frac{1}{2}+\frac{1}{2}(\Pr[D(f(U_n)b(U_n))=1]-\Pr[D(f(U_n)\bar{b}(U_n)]=1)$$
$$=\frac{1}{2}+\frac{1}{2}(\Pr[D(E^{(1)}=1]-\Pr[D(E^{(2)})=1])$$
$$>\frac{1}{2}+\varepsilon.$$ Thus we reach a contradiction.

Let $I=(I_k)_{k \in N}$ be a key set with security parameter k, and let $Q \in Z[X]$ be a positive polynomial. Let $f=(f_i : D_i \to D_i)_{i \in I}$ be a family of one-way permutations with hard core predicate $B=(B_i : D_i \to \{0,1\})_{i \in I}$ and key generator K. Let G=G(f,B,Q) be the induced pseudorandom bit generator.

Is this a PR Bit Generator?



Proof

Then for every P.P.T A with inputs $\mathrm{i} \in \mathrm{I}_\mathrm{k}$, $\mathrm{z} \in \{0,1\}^{\mathrm{Q(k)}}$,
$y \in D_i$ and output in {0,1}:

$$|\Pr(\mathrm{A}(\mathrm{i},\mathrm{G}_\mathrm{i}(x), f_i^{Q(k)}(x)) = 1 : i \leftarrow K(1^k), x \leftarrow D_i)$$
$$- \Pr(A(i,z,y) = 1 : i \leftarrow K(1^k), z \leftarrow \{0,1\}^{Q(k)}, y \leftarrow D_i)| \leq \varepsilon(k)$$

Remark: The theorem states that for sufficiently large keys the probability of distinguishing successfully between truly random sequences and pseudorandom sequences-using a given efficient algorithm is negligibly small, even if $f_i^{Q(k)}(x)$ is known.

Contradicting the pseudo-randomness:

$$\Pr(\mathrm{A}(\mathrm{i},\mathrm{G}_\mathrm{i}(x), f_i^{Q(k)}(x)) = 1 : i \leftarrow K(1^k), x \leftarrow D_i)$$
$$- \Pr(A(i,z,y) = 1 : i \leftarrow K(1^k), z \leftarrow \{0,1\}^{Q(k)}, y \leftarrow D_i) > \varepsilon(k)$$

For $\mathrm{k} \in \mathrm{K}$ and $\mathrm{i} \in \mathrm{I}_\mathrm{k}$, we consider the following sequence of distributions: $p_{i,0}, p_{i,1}, ..., p_{i,Q(k)}$ on $\mathrm{Z}_\mathrm{i} = \{0,1\}^{Q(k)} \times D_i$.

The Hybrid Construction

For $k \in K$ and $i \in I_k$, we consider the following sequence of distributions: $p_{i,0}, p_{i,1}, ..., p_{i,Q(k)}$ on $Z_i = \{0,1\}^{Q(k)} \times D_i$.

$p_{i,0} = \{(b_1, ..., b_{Q(k)}, y) : (b_1, ..., b_{Q(k)}) \leftarrow \{0,1\}^{Q(k)}, y \leftarrow D_i\}$

$p_{i,1} = \{(b_1, ..., b_{Q(k)-1}, B_i(x), f_i(x)) : (b_1, ..., b_{Q(k)-1}) \leftarrow \{0,1\}^{Q(k)-1}, x \leftarrow D_i\}$

...

$p_{i,r} = \{(b_1, ..., b_{Q(k)-r}, B_i(x), B_i(f_i(x)), ..., B_i(f_i^{r-1}(x)), f_i^r(x)) : (b_1, ..., b_{Q(k)-r}) \leftarrow \{0,1\}^{Q(k)-r}, x \leftarrow D_i\}$

...

$p_{i,Q(k)} = \{B_i(x), B_i(f_i(x)), ..., B_i(f_i^{Q(k)-1}(x)), f_i^{Q(k)}(x)) : x \leftarrow D_i\}$

From the contradiction

$\text{Prob}(A(i,z,y)=1; i \leftarrow K(k), z \leftarrow \{0,1\}^{Q(k)}, y \leftarrow D_i)$

$= \text{Prob}(A(i,z,y)=1 : i \leftarrow K(k), (z,y) \xleftarrow{p_{i,0}} Z_i)$

$\text{Prob}(A(i,G_i(x),f_i^{Q(k)}(x))=1; i \leftarrow K(k), z \leftarrow \{0,1\}^{Q(k)}, y \leftarrow D_i)$

$= \text{Prob}(A(i,z,y)=1 : i \leftarrow K(k), (z,y) \xleftarrow{p_{i,Q(k)}} Z_i)$

Thus our contradiction says that algorithm A is able to distinguish between $p_{i,0}$ (uniform distribution) and $p_{i,Q(k)}$ (of pseudorandom sequences).

Difference between each iteration

Since f is bijective,

$p_{i,r} = \{(b_1,...,b_{Q(k)-r}, B_i(x), B_i(f_i(x)),..., B_i(f_i^{r-1}(x)), f_i^r(x)) : (b_1,...,b_{Q(k)-r}) \leftarrow \{0,1\}^{Q(k)-r}, x \leftarrow D_i\}$

$= \{(b_1,...,b_{Q(k)-r}, B_i(f_i(x)), B_i(f_i^2(x)),..., B_i(f_i^r(x)), f_i^{r+1}(x)) : (b_1,...,b_{Q(k)-r}) \leftarrow \{0,1\}^{Q(k)-r}, x \leftarrow D_i\}$

We see that $p_{i,r}$ differs from $p_{i,r+1}$ only at one position, namely at Q(k)-r. There the hard core bit $B_i(x)$ is replaced by a truly random bit.

$$\frac{1}{P(k)} < \text{Prob}(A(i,z,y)=1 : i \leftarrow K(k),(z,y) \xleftarrow{p_{i,Q(k)}} Z_i) - \text{Prob}(A(i,z,y)=1 : i \leftarrow K(k),(z,y) \xleftarrow{p_{i,0}} Z_i)$$

$$= \sum_{r=0}^{Q(k)-1} (\text{Prob}(A(i,z,y)=1 : i \leftarrow K(k),(z,y) \xleftarrow{p_{i,r+1}} Z_i) - \text{Prob}(A(i,z,y)=1 : i \leftarrow K(k),(z,y) \xleftarrow{p_{i,r}} Z_i))$$

Define algorithm A' using A

Choose r, with $0 \le r < Q(k)$, uniformly at random.

Independently choose random bits $b_1, b_2, ..., b_{Q(k)-r-1}$ and another random bit b.

For $y = f_i(x) \in D_i$

$$A'(i, f_i(x)) = \begin{cases} b, \text{ if } A(i, b_1,..,b_{Q(k)-r-1}, b, B_i(f_i(x)),..., B_i(f_i^r(x)), f_i^{r+1}(x)) = 1 \\ 1-b \text{ otherwise} \end{cases}$$

If A distinguishes between $p_{i,r}$ and $p_{i,r+1}$ it yields 1 with higher probability if the (Q(k)-r)th bit of its input is $B_i(x)$ and is not a random bit.

Success of A' in guessing the hard-core predicate

$\Pr(A'(i, f_i(x)) = B_i(x) : i = K(k), x \leftarrow D_i)$

$= \frac{1}{2} + \Pr[A'(i, f_i(x)) = b \mid B_i(x) = b) - \Pr(A'(i, f_i(x)) = b)$

Choosing r uniformly,

$= \frac{1}{2} + \sum_{r=0}^{Q(k)-1} \Pr(R = r).[\Pr(A'(i, f_i(x)) = b \mid B_i(x) = b, R = r) - \Pr(A'(i, f_i(x)) = b \mid R = r)]$

$= \frac{1}{2} + \frac{1}{Q(k)} \sum_{r=0}^{Q(k)-1} [\Pr(A'(i, f_i(x)) = b \mid B_i(x) = b) - \Pr(A'(i, f_i(x)) = b]$

$= \frac{1}{2} + \frac{1}{Q(k)} \sum_{r=0}^{Q(k)-1} (\Pr[A(i, z, y) = 1 : i \leftarrow K(1^k), (z, y) \xleftarrow{p_{i,r+1}} Z_i) -$

$\sum_{r=0}^{Q(k)-1} (\Pr[A(i, z, y) = 1 : i \leftarrow K(1^k), (z, y) \xleftarrow{p_{i,r}} Z_i)$

$> \frac{1}{2} + \frac{1}{Q(k)P(k)}$

This contradicts the hard-core predicate property.

Next Bit Unpredictability

Let $X = (X_1 X_2 ... X_n)$ be a distribution on $\{0,1\}^n$.
X is next-bit unpredictable if for every PPT predictor algorithm P, there exists a negligible function $\varepsilon(n)$ such that,

$$\Pr_{i \in [n]}[P(X_1 ... X_{i-1}) = X_i] \leq \frac{1}{2} + \varepsilon(n)$$

Surprisingly next-bit unpredictability is equivalent to pseudorandomness.

Yao's Theorem

X is pseudorandom if and only if, it is next bit unpredictable.

Proof

X is pseudorandom if and only if, it is next bit unpredictable.

X is PR $\Rightarrow$ Next bit is unpredictable

$\neg$Next bit is unpredictable $\Rightarrow \neg$X is PR

$$\Pr_{i \in_R [n]}[P(X_1 \ldots X_{i-1}) = X_i] > \frac{1}{2} + \varepsilon(n)$$

$$\exists i,\ \Pr[P(X_1 \ldots X_{i-1}) = X_i] > \frac{1}{2} + \varepsilon(n)$$

Define T such that:

$$T(y_1 \ldots y_n) = \begin{cases} 0, & \text{if } P(y_1 \ldots y_{i-1}) \neq y_i \\ 1, & \text{if } P(y_1 \ldots y_{i-1}) = y_i \end{cases}$$

$$\Pr_{y \in U_n}[T(y) = 1] = \frac{1}{2}$$

$$\Pr_{y \in X}[T(y) = 1] > \frac{1}{2} + \varepsilon(n)$$

$Adv(T) > \varepsilon(n)$, thus violating the PRNG property.

Proof of the converse

Let us prove the converse.
Suppose X is not PRNG. Then there is a PPT algorithm T st.:
Adv(T)=|Pr[T(X)=1]-Pr[T(U_n)=1]|>$\varepsilon(n)$
wlog assume Pr[T(X)=1]> Pr[T(U_n)=1].
Now construct a next bit predictor:
Let $U_1,...,U_n$ be uniformly distributed random variables on {0,1}.

$$D_0 = (U_1...U_n)$$
$$D_1 = (X_1...U_n)$$
$$...$$
$$D_{i-1} = (X_1...X_{i-1}U_i...U_n)$$
$$D_i = (X_1...X_iU_{i+1}...U_n)$$
$$...$$
$$D_n = (X_1...X_n)$$

$$\varepsilon(n) < \Pr[T(D_n)=1] - \Pr[T(D_0)=1]$$
$$= \sum_i (\Pr[T(D_i)=1] - \Pr[T(D_{i-1})=1])$$

$\exists i$, st. $\Pr[T(D_i)=1] - \Pr[T(D_{i-1})=1] > \dfrac{\varepsilon(n)}{n}$

Define predictor algorithm $P(x_1...x_{i-1})$:
Choose random bits, $y_i...y_n$.

Let, $P(x_1...x_{i-1}y_i...y_n) = \begin{cases} y_i, if \ T(x_1...x_{i-1}y_i...y_n)=1 \\ 1-y_i, \text{ otherwise} \end{cases}$

Thus, $\Pr[P(X_1...X_{i-1}U_i...U_n) = X_i]$

$$= \frac{1}{2}(\Pr[P(X_1...X_{i-1}U_i...U_n) = X_i \mid U_i = X_i] + \Pr[P(X_1...X_{i-1}U_i...U_n) = X_i \mid U_i = 1 - X_i])$$

$$= \frac{1}{2}(\Pr[P(X_1...X_{i-1}X_i...U_n) = X_i] + \Pr[P(X_1...X_{i-1}1-X_i...U_n) = X_i])$$

$$= \frac{1}{2}(\Pr[T(X_1...X_{i-1}X_i...U_n) = 1] + \Pr[T(X_1...X_{i-1}1-X_i...U_n) = 0])$$

$$= \frac{1}{2}(\Pr[T(D_i) = 1] +$$

$$1 - \Pr[T(X_1 \ldots X_{i-1} 1 - X_i \ldots U_n) = 1])$$

$$= \frac{1}{2} + \frac{1}{2}([\Pr[T(D_i) = 1] - \Pr[T(X_1 \ldots X_{i-1} 1 - X_i \ldots U_n) = 1])$$

$$= \frac{1}{2} + ([\Pr[T(D_i) = 1] - \Pr[T(D_{i-1}) = 1])$$

$$> \frac{1}{2} + \frac{1}{n}(\varepsilon(n))$$

Thus, X is not next bit unpredictable.