Κρυπτογραφία
Ασκήσεις στη Στατιστική Απόσταση

27 Οκτωβρίου 2010

Algorithm 1 Δειγματολήπτης 1

1: Input: A
   Output: A random number $\in [0, \ldots, A)$
2: $n = \lfloor \log_2 A \rfloor + 1$
3: choose $x_0, x_1, \ldots, x_{n-1} \xleftarrow{r} \{0, 1\}$
4: $y = \sum_{i=0}^{n-1} 2^i x_i$
5: return $y \mod A$

Για την ομοιόμορφη κατανομή έχουμε ότι

$$\mathsf{Prob}_{D_1}[u] = \frac{1}{A}, u \in [0, \ldots, A)$$

Παρατηρούμε ότι $2^{n-1} + 1 \leq A < 2^n$ και $2^n \mod A = 2^n - A$, αφού ο δειγματολήπτης σταματά μετά από n βήματα Για την κατανομή που προκύπτει από τον δειγματολήπτη 1 έχουμε

$$\mathsf{Prob}_{D_2}[u] = \begin{cases} \frac{2}{2^n}, & u \in [0, 2^n - A) \\ \frac{1}{2^n} & u \in [2^n - A, A) \end{cases}$$

Θεωρώντας $V = X([D_1]) \cup Y([D_2])$ και $[D_1] = [D_2] = [0, A)$ για την στατι-

στική απόσταση θα έχουμε ότι:

$$\begin{aligned}
\Delta[X,Y] &= \frac{1}{2}\sum_{u\in[0,A)}\left|\underset{X\leftarrow D_1}{\mathsf{Prob}}[X=u]-\underset{X\leftarrow D_2}{\mathsf{Prob}}[X=u]\right| \\
&= \frac{1}{2}\left(\sum_{u\in[0,2^n-A)}\left|\frac{2}{2^n}-\frac{1}{A}\right|+\sum_{u\in[2^n-A,A)}\left|\frac{1}{A}-\frac{1}{2^n}\right|\right) \\
&= \frac{1}{2}\left((2^n-A)\left(\frac{2A-2^n}{A\cdot 2^n}\right)+(2A-2^n)\left(\frac{2^n-A}{A\cdot 2^n}\right)\right) \\
&= \frac{(2^n-A)(2A-2^n)}{A\cdot 2^n}
\end{aligned}$$

□

Algorithm 2 Δειγματολήπτης 2

1: Input: A
   Output: A random number $\in [0, \ldots, A)$
2: $n = \lfloor \log_2 A \rfloor + 1$
3: choose $x_0, x_1, \ldots, x_{n-1} \xleftarrow{r} \{0, 1\}$
4: $y = \sum_{i=0}^{A-1} x_i$
5: return y

Έστω D_1 η ομοιόμορφη κατανομή στο $[0, A)$ και D_2 η διωνυμική κατανομή στο $[0, A)$. Θεωρούμε την τυχαία μεταβλητή X επί του $[0, A)$ και έχουμε:

$$\mathsf{Prob}_{D_1}[X = u] = \frac{1}{A},\ u \in [0, A)$$

Αντίστοιχα για την D_2 θα έχουμε:

$$\mathsf{Prob}_{D_2}[X = u] = \binom{A}{k} \frac{1}{2^k} \frac{1}{2^{A-k}} = \binom{A}{k} \frac{1}{2^A}$$

Βάσει των παραπάνω θα έχουμε ότι:

$$\begin{aligned}
\Delta[X, Y] &= \frac{1}{2} \sum_{u \in [0,A)} \left| \underset{X \leftarrow D_1}{\mathsf{Prob}}[X = u] - \underset{Y \leftarrow D_2}{\mathsf{Prob}}[Y = u] \right| \\
&= \frac{1}{2} \sum_{k=0}^{A-1} \left| \frac{1}{A} - \binom{A-1}{k} \frac{1}{2^A} \right| \\
&= \frac{1}{2} \sum_{k=0}^{A-1} \left| \frac{1}{A} - \frac{(A-1)!}{k!(A-1-k)!2^A} \right|
\end{aligned}$$

□

Για να υπολογίσουμε το παραπάνω άθροισμα, πρέπει να βρούμε τα σημεία στα οποία ισχύει ότι

$$\frac{1}{A} = \binom{A}{k} \frac{1}{2^A}$$

Επειδή το A είναι μεγάλο, μπορούμε να προσεγγίσουμε την διωνυμική κατανομή από την κανονική

$$N(Ap, Ap(1-p))$$

που στην περίπτωση που εξετάζουμε θα είναι η

$$N(A/2, A/4)$$

Επιλύοντας την εξίσωση

$$\frac{2e^{-2\frac{(x-A/2)^2}{A}}}{\sqrt{2\pi A}} = \frac{1}{A}$$

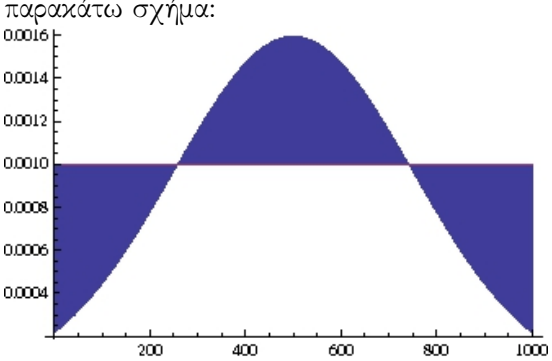
βρίσκουμε ότι τα σημεία τομής της κανονικής και της ομοιόμορφης κατανομής είναι τα

$$x_1 = \frac{1}{2}\left(A - \sqrt{A\log\frac{2A}{\pi}}\right)$$

και

$$x_2 = \frac{1}{2}\left(A + \sqrt{A\log\frac{2A}{\pi}}\right)$$

Έτσι για τον υπολογισμό της στατιστικής απόστασης των δύο κατανομών, αρκεί να υπολογίσουμε το εμβαδόν του γραμμοσκιασμένου τμήματος που φαίνεται στο παρακάτω σχήμα:



Για να διευκολυνθούμε στις πράξεις παρατηρούμε ότι λόγω συμμετρίας αρκεί να υπολογίσουμε το εμβαδόν για $x \in [0, A/2]$. Τότε η στατιστική απόσταση θα είναι το διπλάσιο του εμβαδού αυτού.
Βάσει των παραπάνω θα έχουμε ότι:

$$\begin{aligned}\Delta[X,Y] &= 2\left(\int_0^{x_1}\frac{1}{A} - \frac{e^{-2(x-A/2)^2}\sqrt{2/\pi}}{\sqrt{A}} + \int_{x_1}^{\frac{A}{2}}\frac{e^{-2(x-A/2)^2}\sqrt{2/\pi}}{\sqrt{A}} - \frac{1}{A}\right)\\ &= 2\left(\frac{A + 4x_1 - 3\sqrt{A}\mathrm{Erf}\left(\frac{A}{\sqrt{2}}\right) + 2\sqrt{A}\mathrm{Erf}\left(\frac{A-2x_1}{\sqrt{2}}\right) + 2\sqrt{A}\mathrm{Erf}\left(\sqrt{\frac{\log\frac{2A}{\pi}}{2}}\right) - 2\sqrt{\log\frac{2A}{\pi}}}{2A}\right)\end{aligned}$$

και υπολογίζοντας το όριο της παραπάνω παράστασης καθώς το $A \to +\infty$

έχουμε ότι

$$\lim_{A\to+\infty} 2\left(\frac{A+4x_1-3\sqrt{A}\mathrm{Erf}\left(\frac{A}{\sqrt{2}}\right)+2\sqrt{A}\mathrm{Erf}\left(\frac{A-2x_1}{\sqrt{2}}\right)+2\sqrt{A}\mathrm{Erf}\left(\sqrt{\frac{\log\frac{2A}{\pi}}{2}}\right)-2\sqrt{\log\frac{2A}{\pi}}}{2A}\right)=1$$

Algorithm 3 Δειγματολήπτης 3

```
Input: A
Output: A random number ∈ [0, ..., A)
n = ⌊log_2 A⌋ + 1
choose x_0, x_1, ..., x_{n-1} ←r {0,1}
y = Σ_{i=0}^{n-1} 2^i x_i
if y < A then
    return y
    exit
else
    repeat
end if
```

Έστω D_1 η ομοιόμορφη κατανομή στο $[0, A)$ και D_2 η κατανομή που προκύπτει απο το δειγματολήπτη 3.
Αφού η D_1 είναι η ομοιόμορφη κατανομή, θα έχουμε ότι $\underset{D_1}{\mathsf{Prob}}[u] = 1/A$
Υπολογίζουμε τώρα την πιθανότητα $\underset{D_2}{\mathsf{Prob}}X = u$. Προς τούτο, ορίζουμε δύο τυχαίες μεταβλητές, $X =$ "Ο Δειγματολήπτης εξάγει τον αριθμό u"και $T =$ "Το πείραμα επαναλαμβάνεται t φορές". Έτσι θα έχουμε ότι:

$$\begin{aligned}
\mathsf{Prob}[X=u] &= \sum_{t=1}^{\infty}\mathsf{Prob}[X=u|t]\cdot\mathsf{Prob}[t] \\
&= \frac{1}{A}\left(1-\frac{A}{2^n}\right)\frac{A}{2^n}+\frac{1}{A}\left(1-\frac{A}{2^n}\right)^2\frac{A}{2^n}+\ldots+\frac{1}{2}\left(1-\frac{A}{2^n}\right)^n\frac{A}{2^n}+\ldots \\
&= \frac{1}{2^n}\cdot\frac{2^n}{A} \\
&= \frac{1}{A}
\end{aligned}$$

Βάσει των παραπάνω για τη στατιστική απόσταση έχουμε:

$$
\begin{aligned}
\Delta[X,Y] &= \frac{1}{2}\sum_{u\in[0,2^n)}\left|\underset{X\leftarrow D_1}{\mathsf{Prob}}[X=u]-\underset{X\leftarrow D_2}{\mathsf{Prob}}[X=u]\right| \\
&= \frac{1}{2}\left(\sum_{u\in[0,A)}\left(\frac{1}{A}-\frac{1}{A}\right)\right) \\
&= 0
\end{aligned}
$$

□