Mohamed Aly, Christine Chung (scribe)**, Mahmoud Elhaddad,**
Sherif Khattab, and Jonathan Misurda ("Group A") **Homework 4**
CS 3150 Randomness and Computation **Due:** Wed, Jan 25, 2006

Chapter 5, Exercise 5.10 (page 120)

Consider throwing m balls into n bins, and for convenience let the bins be numbered from 0 to $n-1$. We say there is a *k-gap* starting at bin i if bins $i, i+1, ..., i+k-1$ are empty.

(a) Determine the expected number of k-gaps.

Let the random variable X be the number of k-gaps in the n bins. Let $X_i = 1$ if there exists a k-gap starting at bin i and $X_i = 0$ otherwise. Then, $X = \sum_{i=0}^{n-k} X_i$, and by linearity of expectation,

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{i=0}^{n-k} X_i\right] = \sum_{i=0}^{n-k} \mathbf{E}[X_i]. \tag{1}$$

$$\begin{aligned}
\mathbf{E}[X_i] &= \mathbf{Pr}[X_i = 1] & (2)\\
&= \mathbf{Pr}\left[\bigcap_{b=1}^{m} \text{ball } b \text{ does not land in bins } i \text{ through } i+k-1\right] & (3)\\
&= \prod_{b=1}^{m} \mathbf{Pr}[\text{ball } b \text{ lands in one of the } n-k \text{ other bins}] & (4)\\
&= \left(\frac{n-k}{n}\right)^m. & (5)
\end{aligned}$$

Plugging (5) into (1) gives us

$$\sum_{i=0}^{n-k} \left(\frac{n-k}{n}\right)^m = (n-k+1)\left(\frac{n-k}{n}\right)^m. \tag{6}$$

(b) Prove a Chernoff-like bound for the number of k-gaps. (Hint: If you let $X_i = 1$ when there is a k-gap starting at bin i, then there are dependencies between X_i and X_{i+1}; to avoid these dependencies, you might consider X_i and X_{i+k}.)

Recall from part (a) that we have indicator random variables X_i, where

$$X_i = \begin{cases} 1 & \text{if there is a } k\text{-gap at } i \\ 0 & \text{otherwise} \end{cases}$$

for $0 \le i \le n-k$, and the random variable $X = \sum_{i=0}^{n-k} X_i =$ the number of k-gaps in the n bins after m balls are thrown.

Let the random variable x_i be the number of balls in bin i and let its corresponding independent Poisson random variable be y_i. Similarly, define an independent Poisson random variable Y_i that corresponds to X_i.

Note that X_i can equivalently be defined as a function of $x_i, x_{i+1}, ..., x_{i+k-1}$ as follows:

$$X_i = f(x_i, x_{i+1}, ..., x_{i+k-1}) = \begin{cases} 1 & \text{if } \sum_{j=i}^{i+k-1} x_j = 0 \\ 0 & \text{otherwise} \end{cases}$$

Similarly, Y_i can be defined as $f(y_i, y_{i+1}, ..., y_{i+k-1})$.

Let $Z_i = X_i + X_{i+2k} + \ldots = \sum_{j=0}^{n/k} X_{i+jk}$ and let $Z_i^P = Y_i + Y_{i+k} + Y_{i+2k} + \ldots = \sum_{j=0}^{n/k} Y_{i+jk}$.

With this in mind, we can express the total number of k-gaps as

$$X = \sum_{i=0}^{k-1} Z_i.$$

Since Z_i is a function of X_i, which is a function of x_i, and $\mathbf{E}[Z_i] = \frac{\mathbf{E}[X]}{k}$ by linearity of expectation, we can let

$$g(x_0, ..., x_{n-1}) = \begin{cases} 1 & \text{if } Z_0 \ge (1+\delta)\frac{\mathbf{E}[X]}{k} \\ 0 & \text{otherwise} \end{cases}$$

and similarly,

$$g(y_0, ..., y_{n-1}) = \begin{cases} 1 & \text{if } Z_0^P \ge (1+\delta)\frac{\mathbf{E}[X]}{k} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\mathbf{E}[g(x_0, ..., x_{n-1})] = \mathbf{Pr}\left[Z_0 \ge (1+\delta)\frac{\mathbf{E}[X]}{k}\right]$$

and

$$\mathbf{E}[g(y_0, ..., y_{n-1})] = \mathbf{Pr}\left[Z_0^P \ge (1+\delta)\frac{\mathbf{E}[X]}{k}\right]$$

By theorem 5.7 on page 101 of the text:

$$\mathbf{E}[g(x_0, ..., x_{n-1})] \le e\sqrt{m}\mathbf{E}[g(y_0, ..., y_{n-1})]$$

So

$$\mathbf{Pr}\left[Z_0 \ge (1+\delta)\frac{\mathbf{E}[X]}{k}\right] \le e\sqrt{m}\mathbf{Pr}\left[Z_0^P \ge (1+\delta)\frac{\mathbf{E}[X]}{k}\right].$$

By the Chernoff bound on a sum of independent Poisson trials (theorem 4.4 page 64),

$$\mathbf{Pr}\left[Z_0^P \geq (1+\delta)\frac{\mathbf{E}[X]}{k}\right] \leq e^{-\mathbf{E}[X]\delta^2/3k}.$$

Finally, by observing that if $X \geq (1+\delta)\mathbf{E}[X]$ then there exists an i such that $Z_i \geq (1+\delta)\frac{\mathbf{E}[X]}{k}$, and then applying union bound and substitution, we know:

$$\begin{aligned}
\mathbf{Pr}[X \geq (1+\delta)\mathbf{E}[X]] &\leq \mathbf{Pr}\left[\bigcup_{j=0}^{k-1} Z_i \geq (1+\delta)\frac{\mathbf{E}[X]}{k}\right] && (7)\\
&\leq \sum_{j=0}^{k-1} \mathbf{Pr}\left[Z_i \geq (1+\delta)\frac{\mathbf{E}[X]}{k}\right] && (8)\\
&\leq ke\sqrt{m}\,\mathbf{Pr}\left[Z_i^P \geq (1+\delta)\frac{\mathbf{E}[X]}{k}\right] && (9)\\
&= ke^{1-\mathbf{E}[X]\delta^2/3k}\sqrt{m}. && (10)
\end{aligned}$$