Large deviations in coin-tossing, and an application to routing in networks

Consider a sequence of \( n \) independent tosses of a coin with heads probability \( p \). The expected number of heads is \( \mu = np \). What is the probability that the number of heads deviates a lot from this? Specifically, what is the probability that it lies outside the range \( (1 \pm \delta)\mu \)?

Actually, we’ll be a bit more general and let each coin toss have its own heads probability \( p_i \).

Define \( X_i = \begin{cases} 1 & \text{if the } i\text{th toss is heads;} \\ 0 & \text{otherwise.} \end{cases} \)

Thus the number of heads is \( X = \sum_{i=1}^{n} X_i \).

Clearly \( E(X_i) = p_i \) and \( \text{Var}(X_i) = p_i(1-p_i) \). Hence \( \mu = E(X) = \sum_i p_i \) and, since the \( X_i \) are independent, \( \sigma^2 = \text{Var}(X) = \sum_i \text{Var}(X_i) = \sum_i p_i(1-p_i) \).

Chebyshev’s inequality gives us a bound on the probability that \( X \) deviates from its expectation:

\[
\Pr[|X - \mu| \geq \delta \mu] \leq \frac{\text{Var}(X)}{(\delta \mu)^2} = \frac{\sum_i p_i(1-p_i)}{(\delta \mu)^2} \leq \frac{\mu}{\delta^2 \mu^2} = \frac{1}{\delta^2 \mu^2}.
\]

(*)

This bound, however, is rather weak. A much stronger one is given by Chernoff’s bound, whose proof we omit. (If you are interested, see e.g. Section 4.1 of Motwani & Raghavan.)

This is probably the most widely used inequality in the analysis of sophisticated randomized algorithms today.

**Theorem 1** [Chernoff’s Bound]

1. \( \Pr[X \leq (1 - \delta)\mu] \leq e^{-\delta^2 \mu/2} \) for all \( 0 < \delta < 1 \).

2. \( \Pr[X \geq (1 + \delta)\mu] \leq \begin{cases} e^{-\delta^2 \mu/3} & \text{for all } 0 < \delta < 1; \\ e^{-\delta^2 \mu/(2+\delta)} & \text{for all } \delta \geq 1. \end{cases} \)

Note that we don’t need to consider \( \delta \geq 1 \) in case 1. (Why?)

Putting the two bounds in the theorem together, we get:

\[ \Pr[|X - \mu| \geq \delta \mu] \leq 2e^{-\delta^2 \mu/3} \] for all \( 0 < \delta < 1 \)

(and something of similar form for \( \delta \geq 1 \) in case 2). Compare this with the bound from Chebyshev: it is much stronger, since \( \mu \) appears as a negative exponential rather than just as a denominator. (However, you should remember that Chernoff’s bound applies only to independent coin tosses, while Chebyshev is valid always.)

Section 4.2 from Motwani & Raghavan describes a typical application of Chernoff’s bound. The example problem is that of routing on a hypercube. Chernoff’s bound is used crucially...
in the analysis at the top of page 78. Here, the coin tosses\footnote{Motwani & Raghavan use the term “Poisson trials” for independent coin tosses.} are the r.v.’s $H_{ij}$ (with $i$ fixed and $j = 1, \ldots, N$), so we are interested in the r.v. $X = \sum_{j=1}^{N} H_{ij}$. Now $\mu = E(X) \leq \frac{n}{2}$, and we want a bound on $\Pr[X \geq 6n]$. Clearly the worst case is $\mu = \frac{n}{2}$, so we take $\delta = 11$. Now case 2 of Chernoff’s bound gives us $\Pr[X \geq 6n] \leq e^{-121\mu/13} = e^{-121n/26} < 2^{-6n}$, as claimed by Motwani & Raghavan.

Ex: After understanding the analysis of randomized routing, convince yourself that the analysis would not work using just the Chebyshev bound.

There are several alternative versions of the Chernoff bound. The following one is useful when we do not know the expectation $\mu$.

**Theorem 2 [Chernoff’s Bound – Second Version]**

1. $\Pr[X \leq \mu - \delta n] \leq e^{-2\delta^2 n}$ for all $0 < \delta < \mu$.
2. $\Pr[X \geq \mu + \delta n] \leq e^{-2\delta^2 n}$ for all $0 < \delta < 1$. 