Deviation inequalities

March 14, 2009

Recall that sometimes, we bound probabilities of events instead of computing the probabilities exactly. One example was the union bound, which we used to upper bound the probability of the union of sets, even when the sets are not disjoint.

Here we will try and bound the probability that a random variable deviates from its expectation.

1 Markov and Chebychev’s inequalities

These inequalities play a very important role in characterizing how random variables behave. The Markov inequality holds for any non-negative valued random variable $X$. The Markov inequality for $X$ and all $a > 0$ bounds the probability that $X$ is larger than $a$, specifically

$$P(X > a) \leq \frac{EX}{a}.$$

To derive this inequality, note that

$$EX = \sum_{x \in X(\Omega)} x P(X = x) + \sum_{x \in X(\Omega)} x P(X = x)$$

$$\geq \sum_{x \in X(\Omega)} x P(X = x)$$

$$\geq \sum_{x \in X(\Omega)} a P(X = x)$$

$$= a \sum_{x \in X(\Omega)} P(X = x) = aP(X > a).$$

Equivalently,

$$P(X \leq a) \geq 1 - \frac{EX}{a}.$$

Chebychev’s inequality follows by applying the Markov inequality to the non-negative random variable $g(X) = (X - EX)^2$. It follows that for all $a^2$,

$$E(g(X) \geq a^2) \geq 1 - \frac{Eg(X)}{a^2}.$$

However, $Eg(X) = \text{var}(X)$, so

$$P(g(X) \leq a^2) \geq 1 - \frac{\text{var}(X)}{a^2}. \tag{1}$$

Now what does $g(X) = (X - EX)^2 \leq a^2$ mean? If $a > 0$, the following statements are all true

$$(X - EX)(X - EX) \leq a^2 \tag{2}$$

$$-(X - EX) \cdot -(X - EX) \leq a^2. \tag{3}$$
If \( X \geq EX \), then (2) gives us \( X - EX \geq a \) or \( X \leq EX + a \). If \( X \leq EX \), then (3) gives us \(-(X - EX) \leq a\) or \( X \geq EX - a \). Furthermore, if \( EX - a \leq X \leq EX + a \) then \((X - EX)^2 \leq a^2\). Therefore, we obtain
\[
(X - EX)^2 \leq a^2 \Leftrightarrow EX - a \leq X \leq EX + a.
\]
It follows that
\[
P((X - EX)^2 \leq a^2) = P(EX - a \leq X \leq EX + a),
\]
and hence from (1),
\[
P(EX - a \leq X \leq EX + a) \geq 1 - \frac{\text{var}(X)}{a^2}. \tag{4}
\]
Therefore, the random variable \( X \) is approximately \( EX \) (with an error of \( \pm a \)) with confidence probability at least \( 1 - \frac{\text{var}(X)}{a^2} \). Suppose we let \( a = k\sqrt{\text{var}(X)} \). The quantity \( \sqrt{\text{var}(X)} \) has a special name—standard deviation and is denoted by \( \sigma(X) \). Then,
\[
P(EX - k\sigma(X) \leq X \leq EX + k\sigma(X)) \geq 1 - \frac{1}{k^2}. \tag{5}
\]
So with probability \( 1 - \frac{1}{k^2} \), we can be confident that \( X \) lies in a window of size \( 2k\sigma(X) \) around its expectation. If \( \sigma(X) \) is large, \( X \) is “spread out” quite a bit around its expectation, if \( \sigma(X) \) is small, \( X \) is concentrated around its expectation. Computing \( \sigma \) may or may not be possible. If we have a upper bound \( \sigma^+ \geq \sigma \), then
\[
P(EX - k\sigma^+(X) \leq X \leq EX + k\sigma^+(X)) \geq P(EX - k\sigma(X) \leq X \leq EX + k\sigma(X)) \geq 1 - \frac{1}{k^2}. \tag{5}
\]