Chapter 8

Bounds and Limit Theorems

8.1 Three Probability Bounds

The Markov, Chebychev and Chernoff inequalities provide bounds on the tail of a distribution, i.e., on the probability that a random variable $X$ exceeds some given value — a value that is usually greater than the mean. The Markov inequality, uses only the expectation of a random variable and provides bounds that are often rather loose. The Chebychev inequality, uses both the expectation and the variance of a random variable and usually generates tighter bounds. The Chernoff bound, requires a knowledge of the moment generating function of a random variable. Since a knowledge of a moment generating function implicitly implies a knowledge of moments of all orders, we should expect the Chernoff bound to provide tighter bounds than those obtained from either the Markov or Chebychev inequalities.

8.1.1 The Markov Inequality

Let $X$ be a random variable and $h$ a nondecreasing, nonnegative function. The expectation of $h(X)$, assuming it exists, is given by

$$E[h(X)] = \int_{-\infty}^{\infty} h(u)f_X(u)du,$$

and we may write

$$\int_{-\infty}^{\infty} h(u)f_X(u)du \geq \int_{t}^{\infty} h(u)f_X(u)du \geq h(t) \int_{t}^{\infty} f_X(u)du = h(t)\text{Prob}\{X \geq t\}.$$

This leads directly to the so-called Markov inequality,

$$\text{Prob}\{X \geq t\} \leq \frac{E[h(X)]}{h(t)}.$$  \hspace{1cm} (8.1)

The Markov inequality is exceedingly simple, requiring only the expectation of the distribution. It is used primarily when $t$ is large and $E[X]/t$ is small which means that
a relatively tight bound can be obtained. When these conditions do not hold, the bound can be very loose indeed. Bear in mind also, that this inequality is applicable only to random variables that are nonnegative. The Markov inequality is often applied using $h(x) = x$, in which case it reduces to

$$\Pr\{X \geq t\} \leq \frac{E[X]}{t}, \quad t > 0.$$  

**Example 8.1** Let $X$ be the random variable that denotes the age (in years) of a randomly selected child in Centennial Campus Middle school. If the average child’s age at that school is 12.5 years, then, using the Markov inequality, the probability that a child is at least 20 years old satisfies the inequality

$$\Pr\{X \geq 20\} \leq \frac{12.5}{20} = 0.6250,$$

which is a remarkably loose bound. While this bound is obviously correct, the probability that there is a child aged 20 in a middle school should be very close to zero.

**8.1.2 The Chebychev Inequality**

As we have just seen, the Markov inequality is a first-order inequality in that it requires only knowledge of the mean value. The Chebychev inequality is second order: it requires both the mean value and the variance of the distribution. It may be derived from the Markov inequality as follows. Let the variance $\sigma_X^2$ be finite and define a new random variable $Y$ as

$$Y \equiv (X - E[X])^2.$$  

As in the simple form of the Markov inequality, let $h(x) = x$. Then, from the Markov inequality,

$$\Pr\{Y \geq t^2\} \leq \frac{E[Y]}{t^2}.$$  

Observe that

$$\Pr\{Y \geq t^2\} = \Pr\{(X - E[X])^2 \geq t^2\} = \Pr\{|X - E[X]| \geq t\}$$

and that

$$E[Y] = E[(X - E[X])^2] = \sigma_X^2,$$

from which we immediately have the Chebychev inequality

$$\Pr\{|X - E[X]| \geq t\} \leq \frac{\sigma_X^2}{t^2}, \quad (8.2)$$

From Equation (8.2), it is apparent that the random variable $X$ does not stray far from its mean value $E[X]$ when its variance $\sigma_X^2$ is small. If we set $t = c \sigma_X$, for some positive constant $c$, we obtain

$$\Pr\{|X - E[X]| \geq c \sigma_X\} \leq \frac{1}{c^2}, \quad (8.3)$$
and thus the probability that a random variable is greater than \(c\) standard deviations from its mean is less than \(1/c^2\). Setting \(c = 2\), for example, shows that the probability that a random variable (indeed any random variable) is more than two standard deviations from its mean is less than 1/4.

The Chebychev inequality may be applied to any random variable, unlike the Markov inequality which is applicable only to random variables that are nonnegative. Furthermore, it generally provides a better bound, since it incorporates the variance as well as the expected value of the random variable into the computation of the bound. An alternative form of Equation (8.3) is

\[
\text{Prob}\{|X - E[X]| \leq c\sigma_X\} \geq 1 - \frac{1}{c^2}.
\]

This form of the Chebychev inequality is often used to compute confidence intervals in simulation experiments.

**Example 8.2** Let us return to the same example of middle school children and recompute the bound using the Chebychev inequality. In this case we also need the variance of the ages of the children, which we take to be 3. We seek the probability that a child at the school could be as old as 20. We need to first put this into the form needed by the Chebychev equation.

\[
\text{Prob}\{X \geq 20\} = \text{Prob}\{(X - E[X]) \geq (20 - E[X])\} = \text{Prob}\{(X - E[X]) \geq 7.5\}.
\]

However,

\[
\text{Prob}\{(X - 12.5) \geq 7.5\} \neq \text{Prob}\{|X - 12.5| \geq 7.5\} = \text{Prob}\{5 \leq X \geq 20\},
\]

so that we cannot apply the Chebychev inequality to directly compute \(\text{Prob}\{X \geq 20\}\). Instead, we can compute a bound for \(\text{Prob}\{5 \leq X \geq 20\}\) and obtain

\[
\text{Prob}\{|X - E[X]| \geq 7.5\} \leq \frac{3}{(7.5)^2} = 0.0533,
\]

which is still a much tighter bound than that obtained previously.

### 8.1.3 The Chernoff Bound

The Chernoff bound, just like the Chebychev inequality, may be derived from the Markov inequality. Setting \(h(x) = e^{\theta x}\) for some \(\theta \geq 0\) in Equation (8.1), and using the fact that the moment generating function of a random variable \(X\) is

\[
\mathcal{M}_X(\theta) = E[e^{\theta X}],
\]

we obtain

\[
\text{Prob}\{X \geq t\} \leq \frac{E[e^{\theta X}]}{e^{\theta t}} = e^{-\theta t} \mathcal{M}_X(\theta) \quad \text{for all } \theta \geq 0.
\]