Math 302.102 Fall 2010
Estimates of the Deviation of a Random Variable from its Mean

Our goal is to explain how the standard deviation is a measure of the spread of a distribution.

Theorem (Markov's inequality)**.** *Suppose that Y is a non-negative random variable. If $a > 0$, then*

$$\mathbf{P}\{Y \ge a\} \le \frac{\mathbb{E}(Y)}{a}.$$

Proof. Suppose that $Y \ge 0$ and let $a > 0$. If we define the random variable

$$I = \begin{cases} 1, & \text{if } Y \ge a, \\ 0, & \text{if } Y < a, \end{cases}$$

then $Y \ge aI$ and so $\mathbb{E}(Y) \ge a\mathbb{E}(I)$. However,

$$\mathbb{E}(I) = 1 \cdot \mathbf{P}\{I = 1\} + 0 \cdot \mathbf{P}\{I = 0\} = 1 \cdot \mathbf{P}\{Y \ge a\} + 0 \cdot \mathbf{P}\{Y < a\} = \mathbf{P}\{Y \ge a\}$$

implying that

$$\mathbb{E}(Y) \ge a\mathbf{P}\{Y \ge a\} \quad \text{or, equivalently,} \quad \mathbf{P}\{Y \ge a\} \le \frac{\mathbb{E}(Y)}{a}$$

as required. □

Two special cases of Markov's inequality are often distinguished. Observe that if $Y \ge 0$, then $Y^2 \ge a^2$ if and only if $Y \ge a$. This implies that

$$\mathbf{P}\{Y \ge a\} = \mathbf{P}\left\{Y^2 \ge a^2\right\} \le \frac{\mathbb{E}(Y^2)}{a^2}. \tag{$*$}$$

This is sometimes known as *Chebychev's inequality*. Similarly, $Y \ge a$ if and only if $e^{tY} \ge e^{ta}$ for any $t > 0$ implying that

$$\mathbf{P}\{Y \ge a\} = \mathbf{P}\left\{e^{tY} \ge e^{ta}\right\} \le \frac{\mathbb{E}(e^{tY})}{e^{ta}}.$$

Since $m(t) = \mathbb{E}(e^{tY})$ is the moment generating function of Y, we can rephrase this as

$$\mathbf{P}\{Y \ge a\} \le e^{-ta}m(t) \text{ for } t > 0.$$

This is sometimes known as *Chernoff's inequality.*

Of course, we are often interested in random variables other than those that are non-negative. The general form of Chebychev's inequality is as follows.

Theorem (Chebychev's inequality)**.** *If X is a random variable with mean $\mathbb{E}(X)$ and variance $\mathrm{Var}(X)$, then*

$$\mathbf{P}\{|X - \mathbb{E}(X)| \ge a\} \le \frac{\mathrm{Var}(X)}{a^2}$$

for any $a > 0$.

Proof. Let $Y = |X - \mathbb{E}(X)|$ so that $Y \geq 0$ and apply Chebychev's inequality in the form of $(*)$ to obtain

$$\mathbf{P}\{Y \geq a\} = \mathbf{P}\{|X - \mathbb{E}(X)| \geq a\} \leq \frac{\mathbb{E}(|X - \mathbb{E}(X)|^2)}{a^2} = \frac{\text{Var}(X)}{a^2}$$

using the fact that $\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X)]^2) = \mathbb{E}(|X - \mathbb{E}(X)|^2)$ by definition. □

As an application, if we take $a = 2\,\text{SD}(X)$, then we find

$$\mathbf{P}\{|X - \mathbb{E}(X)| \geq 2\,\text{SD}(X)\} \leq \frac{\text{Var}(X)}{(2\,\text{SD}(X))^2} = \frac{1}{4},$$

and if we take $a = 3\,\text{SD}(X)$, then we find

$$\mathbf{P}\{|X - \mathbb{E}(X)| \geq 3\,\text{SD}(X)\} \leq \frac{\text{Var}(X)}{(3\,\text{SD}(X))^2} = \frac{1}{9}.$$

By taking complements, we can re-write these inequalities as

$$\mathbf{P}\{|X - \mathbb{E}(X)| \leq 2\,\text{SD}(X)\} \geq \frac{3}{4}$$

and

$$\mathbf{P}\{|X - \mathbb{E}(X)| \leq 3\,\text{SD}(X)\} \geq \frac{8}{9}.$$

The interpretation is that if X is *any* continuous random variable with density function f, then at least 75% of the area under f falls within 2 standard deviation of the mean and at least 89% of the area under f falls within 3 standard deviations of the mean. That is,

$$\mathbf{P}\{|X - \mu| \leq 2\sigma\} \geq \frac{3}{4} \quad \text{and} \quad \mathbf{P}\{|X - \mu| \leq 3\sigma\} \geq \frac{8}{9}.$$

Equivalently, we can interpret this result statistically: for any random sample of data consisting of observations that were taken independently, at least 75% of data is within 2 standard deviations of the mean and at least 89% of data is within 3 standard deviations of the mean.

In fact, this is how Chebychev's inequality is usually presented in elementary statistics textbooks. If $k > 0$, then

$$\mathbf{P}\{|X - \mu| \leq k\sigma\} \geq 1 - \frac{1}{k^2}.$$

Exercise. Suppose that $X \sim \text{Exp}(\lambda)$. Compute $\mu = \mathbb{E}(X)$ and $\sigma = \text{SD}(X)$. Sketch a graph of $f(x)$ and mark the points μ, $\mu + 2\sigma$, and $\mu - 2\sigma$ on the horizontal axis. Compute $\mathbf{P}\{|X - \mu| \leq 2\sigma\} = \mathbf{P}\{\mu - 2\sigma \leq X \leq \mu + 2\sigma\}$. How does this compare to the estimate promised by Chebychev's inequality?