Theorem 1 (Chebyshev's Inequality)**.** *Let X be a random variable with density function $f(x)$, mean μ and variance σ^2 and let $k > 1$. Then*

$$\Pr(|X - \mu| > k\sigma) \leq \frac{1}{k^2}.$$

Proof.

$$\begin{aligned}
\Pr(|X - \mu| > k\sigma) &= \int_{\{x:|x-\mu|>k\sigma\}} f(x)\,dx \\
&\leq \int_{\{x:|x-\mu|>k\sigma\}} \frac{(x-\mu)^2}{(k\sigma)^2} f(x)\,dx \\
&\leq \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{(k\sigma)^2} f(x)\,dx \\
&= \frac{1}{(k\sigma)^2} \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)\,dx \\
&= \frac{\sigma^2}{(k\sigma)^2} = \frac{1}{k^2}.
\end{aligned}$$

□

Visualization.

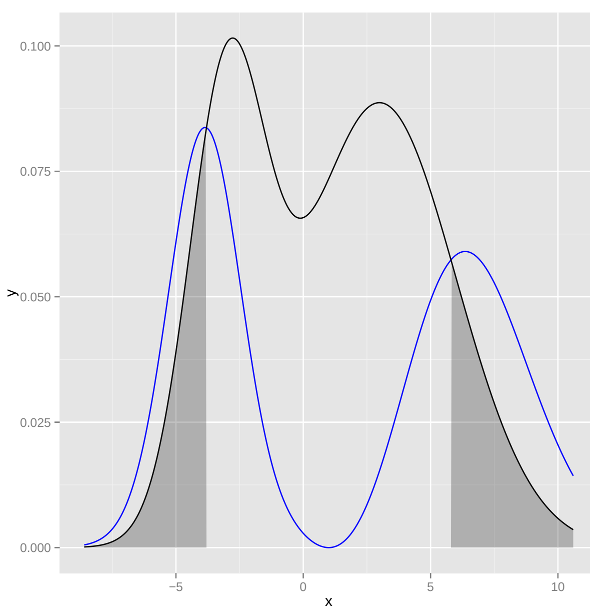

Figure 1: Chebyshev's Inequality Visualization

The black curve in figure 1 is some arbitrary density function, the function $f(x)$ from the proof. The two gray areas are the areas under the density curve that are more than k standard deviations from the mean of the density curve. So the area of the two pieces together is $\Pr(|X - \mu| > k\sigma)$. We want to show that this is less than $1/k^2$.

The blue curve is $f(x)$ times $(x - \mu)^2/(k\sigma)^2$. You can see that the area under the blue curve is more than the area in the two gray pieces. This is what the proof shows, in two steps. The first inequality in the proof says that the area under the blue curve on the left and right tails is more than the area in the gray pieces, because in the area of integration $(x - \mu)^2/(k\sigma)^2$ is always larger than 1. The second inequality just says that if you add the middle part of the area under the blue curve, you get even more.

The key to the proof is that the blue curve was constructed so that the area under it is $1/k^2$. So the area represented by the two gray pieces is less than $1/k^2$.

The figure is produced by this code.

```
# Tuning parameters for the example
# Example function is created by mixing two normal density functions
mean1 <- -3  # mean of first density
sigma1 <- 1.5  # sd of first density
mean2 <- 3   # mean of second density
sigma2 <- 3  # sd of second density
a <- 1/3     # "mixing": 0 <= a <= 1
k <- 1.25     # value of k: k > 1
# Computed values
meanf <- a * mean1 + (1-a) * mean2 # mean of example function
var1 <- sigma1^2  # variance of first density
var2 <- sigma2^2  # variance of second density
expsq1 <- var1 + mean1^2   # E(X_1^2)
expsq2 <- var2 + mean2^2   # E(X_2^2)
expsqf <- a * expsq1 + (1-a) * expsq2   # E(X_f^2)
varf <- expsqf - meanf^2   # variance of f
sigmaf <- sqrt(varf)   # sd of example function
xlower <- meanf - 2 * k * sigmaf # lower limit for graph
xupper <- meanf + 2 * k * sigmaf # upper limit for graph
x <- seq(xlower, xupper, length.out=500)
fx <- a * dnorm(x, mean1, sigma1) + (1-a) * dnorm(x, mean2, sigma2)
# Limits for Chebyshev inequality
limlower <- meanf - k * sigmaf
limupper <- meanf + k * sigmaf
# Add polygons for P(|X-mean| > k * sigma)
xlowertail <- x[x < limlower]
xuppertail <- x[x > limupper]
fxlowertail <- fx[x < limlower]
fxuppertail <- fx[x > limupper]
# Add Chebyshev function
cheb <- fx * ((x - meanf) / (k * sigmaf))^2
# ggplot2 graphs
lines.df <- data.frame(x=rep(x, 2),
                       y=c(fx, cheb),
                       cv=rep(c("density", "Chebyshev"),
                              each=length(x)))
polys.df <- data.frame(x=c(xlowertail, limlower, xlower,
                           xuppertail, xupper, limupper),
                       y=c(fxlowertail, 0, 0,
                           fxuppertail, 0, 0),
                       tail=c(rep("lower", length(xlowertail) + 2),
                              rep("upper", length(xuppertail) + 2)))
ggplot(data=polys.df, aes(x=x,y=y)) +
    geom_polygon(aes(group=tail),alpha=.3) +
    geom_line(data=lines.df,aes(x=x,y=y,colour=cv)) +
    scale_colour_manual(values=c("blue", "black")) +
    guides(colour=FALSE)
```