Maximum distance between the mode and the mean of a unimodal distribution

Henry Bottomley
5 Leydon Close, London SE16 5PF
se16@btinternet.com

6th February 2004

Abstract

The difference between the mean and the mode of a unimodal distribution is less than or equal to the square root of three times the standard deviation.

It is reasonably well known that the difference between the mean and the median of the distribution of a random variable is less than or equal to one standard deviation - providing that all three statistics exist. Oliver Johnson and Yuri Sukov[1] suggest that the similar result for the difference between the mean and the mode is not generally known: in the case of a unimodal distribution, it is in fact less than or equal to $\sqrt{3}$ standard deviations. This note provides a simple proof. The result holds both for continuous and for discrete random variables, providing that *unimodal* is defined sensibly.

For a continuous random variable X with a unimodal probability density function $f_X(x)$, there will be a point which we will call the *mode* and label as $\hat{X}$, such that $f_X(x)$ is weakly increasing below $\hat{X}$ and weakly decreasing above $\hat{X}$. The mode may not be uniquely defined if the maximum density is achieved over an interval, such as the mode of a uniform distribution. In such cases, we would have some limited discretion over which point we choose to be the mode, and we shall use this discretion later. Note that while the random variable X needs to be continuous, the density does not.

For a discrete random variable D, the definition of unimodal is similar but slightly more complicated. Simply having one point having a higher probability than each of the others would be no more satisfactory than doing the same for the density of a continuous random variable; so we need to apply the idea of probability increasing up to the mode and decreasing above it.

The problem is that there will be many intermediate points of zero probability, raising the question of which of these would destroy the unimodal nature of the probability distribution. The answer used here is to require the probability distribution to be restricted to a set of equally spaced points, so if the gap between these points is g, then the entire probability is at points at integer multiples of g from the mode $\hat{D}$, and the probability mass function $\mathbb{P}(D = \hat{D} + ig)$ for integer i is weakly increasing below $\hat{D}$ and is weakly decreasing above $\hat{D}$. Again, we may have some limited discretion over which point we choose to be the mode.

We use the following intuitive lemma to minimise variance. In essence, it says that for two continuous random variables with the same mean, if the density of the first is less than that of the second inside a certain interval, and the density of the first exceeds that of the second outside the interval, then the first has a larger variance than the second.

Lemma 1 *If there are two continuous random variables Y and Z with densities $f_Y(x)$ and $f_Z(x)$, with the same mean μ, and there is an interval (a,b) such that $f_Y(x) \leq f_Z(x)\,\forall x \in (a,b)$ and $f_Y(x) \geq f_Z(x)\,\forall x \in (-\infty,a) \cup (b,\infty)$, then $\sigma_Y^2 \geq \sigma_Z^2$.*

Proof The two densities must each integrate to 1, so

$$\int_{-\infty}^{a} (f_Y(x) - f_Z(x))\mathrm{d}x - \int_{a}^{b} (f_Z(x) - f_Y(x))\mathrm{d}x + \int_{b}^{\infty} (f_Y(x) - f_Z(x))\mathrm{d}x = 0$$

and each of the integrands is non-negative in the range of its integral, so

$$\int_{-\infty}^{a} \left(x - \frac{a+b}{2}\right)^2 (f_Y(x) - f_Z(x))\,\mathrm{d}x \geq \left(\frac{b-a}{2}\right)^2 \int_{-\infty}^{a} (f_Y(x) - f_Z(x))\,\mathrm{d}x$$

$$\int_{a}^{b} \left(x - \frac{a+b}{2}\right)^2 (f_Z(x) - f_Y(x))\,\mathrm{d}x \leq \left(\frac{b-a}{2}\right)^2 \int_{a}^{b} (f_Z(x) - f_Y(x))\,\mathrm{d}x$$

$$\int_{b}^{\infty} \left(x - \frac{a+b}{2}\right)^2 (f_Y(x) - f_Z(x))\,\mathrm{d}x \geq \left(\frac{b-a}{2}\right)^2 \int_{b}^{\infty} (f_Y(x) - f_Z(x))\,\mathrm{d}x$$

and by taking second moments about the point $(a+b)/2$, we get

$$
\begin{aligned}
\sigma_Y^2 - \sigma_Z^2 = & \left(\int_{-\infty}^{\infty} \left(x - \frac{a+b}{2}\right)^2 f_Y(x)\mathrm{d}x - \left(\mu - \frac{a+b}{2}\right)^2 \right) \\
& - \left(\int_{-\infty}^{\infty} \left(x - \frac{a+b}{2}\right)^2 f_Z(x)\mathrm{d}x - \left(\mu - \frac{a+b}{2}\right)^2 \right) \\
= & \int_{-\infty}^{a} \left(x - \frac{a+b}{2}\right)^2 (f_Y(x) - f_Z(x))\mathrm{d}x \\
& - \int_{a}^{b} \left(x - \frac{a+b}{2}\right)^2 (f_Z(x) - f_Y(x))\mathrm{d}x \\
& + \int_{b}^{\infty} \left(x - \frac{a+b}{2}\right)^2 (f_Y(x) - f_Z(x))\mathrm{d}x \\
\geq & \left(\frac{b-a}{2}\right)^2 \left(\int_{-\infty}^{a} (f_Y(x) - f_Z(x))\mathrm{d}x \right. \\
& \left. - \int_{a}^{b} (f_Z(x) - f_Y(x))\mathrm{d}x + \int_{b}^{\infty} (f_Y(x) - f_Z(x))\mathrm{d}x \right) \\
= & \; 0. \quad \blacksquare
\end{aligned}
$$

The proof of the next inequality shown here depends on finding another continuous random variable which meets the definition, has the same mean, and which has a lower variance and so a lower standard deviation.

Theorem 2 *Any continuous random variable X with a unimodal probability density function, with a finite mean μ_X and standard deviation σ_X, and with a mode $\hat{X}$, has:*

$$|\hat{X} - \mu_X| \leq \sqrt{3}\sigma_X.$$

Proof If $\hat{X} = \mu_X$ then we need go no further. Otherwise we will assume $\hat{X} > \mu_X$, since if $\hat{X} < \mu_X$ we can look at $-X$.

We can then compare the distribution of X with an random variable U uniformly distributed between $2\mu_X - \hat{X}$ and $\hat{X}$. If U has the density $f_U(x) = 1/(2\hat{X} - 2\mu_X)$ if $2\mu_X - \hat{X} < x \leq \hat{X}$ and $f_U(x) = 0$ if $x \leq \hat{X}$ or $2\mu_X - \hat{X} < x$, then it has mean $\mu_U = \mu_X$ and variance $\sigma_U^2 = (\hat{X} - \mu_X)^2/3$. We can choose the mode $\hat{U}$ to be $\hat{X}$.

Since $f_X(x)$ is weakly increasing below $\hat{X}$, there is a point $c \in [2\mu_X - \hat{X}, \hat{X}]$ where $f_X(x) \leq f_U(x) \forall x \in (2\mu_X - \hat{X}, c)$ and $f_X(x) > f_U(x) \forall x \in (c, \hat{X})$; c might need to be $\hat{X}$. In any case $f_X(x) \geq 0 = f_U(x)\, \forall x \in (-\infty, 2\mu_X - \hat{X}) \cup (\hat{X}, \infty)$. So by taking the second moment about $\mu_X - (\hat{X} - c)/2$, lemma 1 gives $\sigma_X^2 \geq \sigma_U^2 = (\hat{X} - \mu_X)^2/3$, leading to the result in the theorem. ∎

Not surprisingly, the inequality becomes an equality for a uniform distribution where the mode is taken as being at one end.

The result of lemma 1 can easily be extended to include points $\{x_i\}$ of positive probability where $\mathbb{P}(Y = x_i) \leq \mathbb{P}(Z = x_i)$ inside (a, b) and where $\mathbb{P}(Y = x_i) \geq \mathbb{P}(Z = x_i)$ in $(-\infty, 2\mu_X - \hat{X}) \cup (\hat{X}, \infty)$. Theorem 2 can be extended to cover a random variable X with a point of positive probability at $\hat{X}$, i.e. $\mathbb{P}(X = \hat{X})$ could be positive, but there cannot be any other such points if X is to be basically continuous and have a unimodal distribution.

The inequality in theorem 2 also applies to discrete random variables; the inequality becomes strict providing that the standard deviation is positive.

Theorem 3 *Any discrete random variable D with an equally spaced unimodal probability mass function, with a finite mean μ_D and positive standard deviation σ_D, and with a mode $\hat{D}$, has:*

$$|\hat{D} - \mu_D| < \sqrt{3}\sigma_D.$$

Proof If $\hat{D} = \mu_D$ then we need go no further. Otherwise we will assume $\hat{D} > \mu_D$, since if $\hat{D} < \mu_D$ we can look at $-D$.

We need a suitable continuous random variable X with a unimodal distribution. Fortunately, there is an obvious distribution which with one careful choice will produce the desired result. If the gap for D is $g > 0$ then we can then specify the density of X as

$$f_X(x) = \frac{\mathbb{P}(D = \hat{D} + ig)}{g} \text{ when } \hat{D} + (i - \tfrac{1}{2})g < x \leq \hat{D} + (i + \tfrac{1}{2})g$$

for integer i. This will have mean $\mu_X = \mu_D$, and variance $\sigma_X^2 = \sigma_D^2 + \frac{g^2}{12}$.

Since the distribution of X is made up of parts of uniform density of which the highest is around $\hat{D}$, we have some discretion about selecting a mode for X. Perhaps surprisingly, we will choose $\hat{X} = \hat{D} + \frac{g}{2}$ which trivially gives

$$\hat{X} - \mu_X = \hat{D} - \mu_D + \tfrac{g}{2} > \tfrac{g}{2}.$$

X is a continuous random variable with a unimodal probability distribution, so from theorem 2 we have $(\hat{X} - \mu_X)^2 \leq 3\sigma_X^2$, and

$$\begin{aligned}(\hat{D} - \mu_D)^2 &= (\hat{X} - \mu_X - \tfrac{g}{2})^2 = (\hat{X} - \mu_X)^2 - g(\hat{X} - \mu_X) + \tfrac{g^2}{4} \\ &< (\hat{X} - \mu_X)^2 - \tfrac{g^2}{2} + \tfrac{g^2}{4} \leq 3\sigma_X^2 - 3\tfrac{g^2}{12} = 3\sigma_D^2\end{aligned}$$

and taking the square root gives the result in the theorem. ∎

It is slightly surprising that the inequalities in theorems 2 and 3 are essentially the same. Usually such inequalities involving location and dispersion statistics are tighter for continuous unimodal distributions than for equally spaced discrete unimodal distributions: this is because there are sequences of discrete unimodal distributions which can converge in distribution to any given continuous unimodal distribution by narrowing the gaps, but sequences of continuous distributions which converge in distribution to a given discrete unimodal distribution cannot all be unimodal, because of the gaps.

Carl Friedrich Gauss[2] produced an inequality broadly similar to Chebyshev's inequality, but instead based on the mode and on the second moment about the mode $\mathbb{E}(|X - \hat{X}|^2)$. He called the square root of this *error medius metuendus* in Latin and labelled it m, though something like *rms mode deviation* and $s_{\hat{X}}$ might be clearer (rms for root mean square). Theorems 2 and 3 lead to simple constraints on this deviation.

Theorem 4 *A random variable X (whether continuous or equally spaced discrete) with a unimodal distribution, with a finite mean μ_X and standard deviation σ_X, and with a mode $\hat{X}$ and rms mode deviation $s_{\hat{X}}$, has:*

$$\sigma_X \leq s_{\hat{X}} \leq 2\sigma_X \text{ and } |\hat{X} - \mu_X| \leq \sqrt{\frac{3}{4}} s_{\hat{X}}.$$

Proof We have $s_{\hat{X}}^2 = \sigma_X^2 + (\hat{X} - \mu_X)^2$, as with all second moments.

This gives $\sigma_X^2 \leq s_{\hat{X}}^2$ immediately; since theorems 2 and 3 amount to $(\hat{X} - \mu_X)^2 \leq 3\sigma_X^2$, it also gives $s_{\hat{X}}^2 \leq 4\sigma_X^2$, i.e. $\sigma_X^2 \geq \frac{1}{4}s_{\hat{X}}^2$, and thus it gives $(\hat{X} - \mu_X)^2 \leq \frac{3}{4}s_{\hat{X}}^2$. Taking square roots gives the results in the theorem. ∎

References

[1] O Johnson and Y Suhov, *The von Neumann entropy and information rate for integrable quantum Gibbs ensembles. 2*, Quantum Computers and Computing, Vol 4/1, 2003, pages 128–143.

[2] C F Gauss, *Theoria Combinationis Observationum Erroribus Minimum Obnoxiae*, Royal Society of Gottingen, 1821, pars prior, §§7–10.