Justification for the Sample Size Calculation for Proportions from Desired Margin of Error

We have found that, to estimate the sample size (***n***) needed to estimate the value of $\hat{p}$, we can use one of the following two formulae:

a) $n = \dfrac{\left(z_{\alpha/2}\right)^2 \hat{p}\hat{q}}{E^2}$, when a prior value of $\hat{p}$ is known b) $n = \dfrac{\left(z_{\alpha/2}\right)^2}{E^2}(0.25)$, otherwise

They both are derived from the error calculation

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

but it might not be entirely clear as to why the 0.25 occurs in formula b). Let's find out:

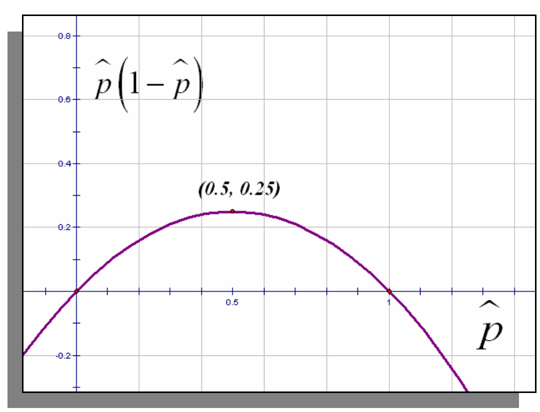
* * * * * * *

When creating confidence intervals, we aim to make the interval as wide as possible (to ensure capturing the true parameter – in this case, a population proportion). Of course, we also want to be limited by practicality, so we aim to make the interval as wide as possible *within certain constraints*. For proportions, the constraints are 1) the confidence level (α) and the value of the sample proportion ($\hat{p}$). To construct a proper sample size, these same two constraints are necessary.

However, if we don't have a value for $\hat{p}$, yet still want to find a proper sample size, we need to substitute a value for $\hat{p}$ to accommodate our needs. To be on the safe side, we should use a value for $\hat{p}$ that maximizes the sample size (and also the confidence interval). Look again at the formula for sample size for a known value of $\hat{p}$:

$$n = \frac{\left(z_{\alpha/2}\right)^2 \hat{p}\hat{q}}{E^2}$$

The key to the 0.25 is in the $\hat{p}\hat{q}$ factor of that formula. We want to find values of $\hat{p}$ and $\hat{q}$ that maximize the value of the product $\hat{p}\hat{q}$. The key to this is recalling $\hat{q} = 1 - \hat{p}$, so $\hat{p}\hat{q} = \hat{p}\left(1-\hat{p}\right)$. Thus, we have to maximize the function given by the rule $\hat{p}\hat{q} = \hat{p}\left(1-\hat{p}\right)$. Looking at the graph of this parabola (a-HA!) we find that the maximum occurs when $\hat{p} = 0.5$, which means that $\hat{q} = 0.5$ as well[1]. Thus, the maximize ***n***,



$$n = \frac{\left(z_{\alpha/2}\right)^2 \hat{p}\hat{q}}{E^2} = \frac{\left(z_{\alpha/2}\right)^2 (0.5)(0.5)}{E^2} = \frac{\left(z_{\alpha/2}\right)^2}{E^2}(0.25)$$

---

[1] If you have taken calculus, the maximum can be found using derivatives: if $f\left(\hat{p}\right) = \hat{p}\left(1-\hat{p}\right)$, then $f'\left(\hat{p}\right) = 1 - 2\hat{p}$. Setting $f\left(\hat{p}\right) = 0$ and solving for $\hat{p}$ gives the desired result.