The Chernoff bound gives a large deviation result for i.i.d. random variables, saying that the probability of the empirical mean to be far from mean itself is exponentially small in the number of random variables. What follows is adapted from Leonard Schulman's lecture course on probability and algorithms at Caltech: http://www.cs.caltech.edu/ schulman/Courses/03cs150/03cs150.html (lecture 4). (If anyone knows the standard reference, pass it along.)

Consider a set of independent random variables $X_1, X_2, \ldots, X_n$ with empirical mean $Y = \frac{1}{n}\sum_{i=1}^n X_i$. Without loss of generality we can assume $\langle X\rangle = 0$. Then provided the characteristic function $\varphi_X(\beta) = \langle e^{\beta X}\rangle$ converges unconditionally in a neighborhood of zero and is differentiable at zero, for all $\epsilon > 0$ there exists $c_\epsilon < 1$ such that

$$p(Y > \epsilon) < c_\epsilon^n.$$

Proof: Use the Bernstein trick and consider the exponential of βY instead of Y itself and figure out what to do with β later. $p(Y > \epsilon) = p(e^{\beta n Y} > e^{\beta n\epsilon})$ for any $\beta > 0$, and by the Markov inequality we have

$$p(e^{\beta nY} > e^{\beta n\epsilon}) < e^{-\beta n\epsilon}\langle e^{\beta nY}\rangle = e^{-\beta n\epsilon}\langle e^{\beta n\sum_{i=1}^n X_i}\rangle = \left(e^{-\beta\epsilon}\langle e^{\beta X}\rangle\right)^n = \left(e^{-\beta\epsilon}\varphi_X(\beta)\right)^n.$$

Now we need to show there's a $\beta > 0$ such that $e^{-\beta\epsilon}\varphi_X(\beta) < 1$. For $\beta = 0$ we get $\varphi(0) = 1$, and the derivative at zero gives

$$\partial_\beta e^{-\beta\epsilon}\varphi_X(\beta)\big|_0 = -\epsilon + \varphi_X'(0).$$

For $\varphi_X'(0)$ the unconditional convergence implies that we can switch the order of expectation value taking (integration) and differentiation, so

$$\varphi_X'(0) = \partial_\beta\langle e^{\beta X}\rangle = \langle\partial_\beta e^{\beta X}\rangle = \langle X e^{\beta X}\rangle = \langle X\rangle = 0.$$

Thus, $\partial_\beta e^{-\beta\epsilon}\varphi_X(\beta)\big|_0 = -\epsilon$ and so there exists a c_ϵ as stated.

Clearly any bounded discrete random variable has a smooth characteristic function, and thus the Chernoff bound applies.

Example: Suppose $X \in \pm 1$ with equal probability. The characteristic function is thus $\varphi_X(\beta) = \frac{1}{2}(e^\beta + e^{-\beta}) = \cosh\beta$. Then $c_\epsilon(\beta) = e^{-\epsilon\beta}\cosh\beta$ which is less than 1 for any $\beta > 0$ as proven above. The best value, $c_\epsilon = \inf_\beta c_\epsilon(\beta)$, occurs for $\beta = \frac{1}{2}\log\frac{1+\epsilon}{1-\epsilon}$, and gives $-\log_2 c_\epsilon = \frac{1-\epsilon}{2}\log_2(1-\epsilon) + \frac{1+\epsilon}{2}\log_2(1-\epsilon)$, which is the relative entropy $D(\{\frac{1+\epsilon}{2}, \frac{1-\epsilon}{2}\}||\{\frac{1}{2}, \frac{1}{2}\}) = r_\epsilon$. Thus we have

$$p(Y > \epsilon) < 2^{-nr_\epsilon}.$$