

AMS 131: Quiz 9

Name: _____

You're about to take an IID sample (X_1, \dots, X_n) from a distribution with variance $V(X_i) = \sigma^2$ that exists and is finite, which implies that the mean $E(X_i) = \mu$ also exists and is finite. The purpose of the sampling is to use the sample mean $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ as an estimator of μ , and you're wondering what value you should use for the sample size n . We've seen in class that *Chebyshev's Inequality* can help when no other details about the distribution of the X_i are available: if Y is any random variable whose variance $V(Y)$ exists, this inequality states that for any $t \geq 0$

$$P(|Y - E(Y)| \geq t) \leq \frac{V(Y)}{t^2}. \quad (1)$$

- (a) Using basic facts about $E(\bar{X}_n)$ and $V(\bar{X}_n)$, show that inequality (1) implies, in the random sampling problem considered here, that for any $k > 0$

$$P(|\bar{X}_n - \mu| < k\sigma) \geq 1 - \frac{1}{nk^2}, \quad (2)$$

and show further that, if we want the probability in (2) to be at least $(1 - \alpha)$ for some $0 < \alpha < 1$, we should choose

$$n_{Chebyshev} \geq \frac{1}{\alpha k^2}. \quad (3)$$

I also mentioned in class that Chebyshev's Inequality can be quite conservative. Let's quantify this: suppose for the rest of the problem that $(X_i | \mu, \sigma^2) \stackrel{IID}{\sim} N(\mu, \sigma^2)$. Then we've seen in class that \bar{X}_n also follows a Normal distribution, with mean μ and standard error $SE(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$. As usual let $\Phi(x)$ be the standard Normal CDF; in other words, $\Phi(x)$ is the area to the left of x under the standard Normal PDF.

- (b) Sketch the PDF of \bar{X}_n , shading in the area corresponding to $(*) = (|\bar{X}_n - \mu| < k\sigma)$ and identifying the places on both the raw-units and standard-units axes corresponding to the endpoints of $(*)$.

- (c) Using a basic fact about $\Phi(x)$ and your sketch in part (b), show that under the Normality assumption for the X_i ,

$$P(|\bar{X}_n - \mu| < k\sigma) \geq 2\Phi(k\sqrt{n}) - 1. \quad (4)$$

- (d) Set the probability on the right-hand side of the inequality in (4) equal to $(1 - \alpha)$ and solve for n , thereby showing that under the Normality assumption the required sample size corresponding to $n_{Chebyshev}$ in equation (3) above is

$$n_{Normality} \geq \frac{[\Phi^{-1}(1 - \frac{\alpha}{2})]^2}{k^2}, \quad (5)$$

in which (as usual) $\Phi^{-1}(p)$ (for $0 < p < 1$) is the inverse CDF (quantile function) for the standard Normal distribution.

- (e) Using the table on page 861 of Degroot and Schervish or (better) an online inverse Normal CDF calculator (e.g., there's one provided by [Wolfram Alpha](#)), complete the rest of the table below.

α	$\frac{1}{\alpha}$	$[\Phi^{-1}(1 - \frac{\alpha}{2})]^2$
0.1	10	2.7
0.05	20	
0.01		6.6
0.005		
0.001	1000	

- (f) If the data values X_i really did come from a Normal distribution, would you describe the Chebyshev Inequality sample size calculation as highly conservative, not too conservative, or in between? Explain briefly.