# All of Statistics - Chapter 4 Solutions

May 9, 2020

## 1.

Chebyshev's inequality gives $\mathbb{P}(|X-\mu| \geq k\sigma) \leq 1/k^2$. An exact calculation yields instead $e^{-(1+k)}$. To see this, note that $\beta(\mu \pm k\sigma) = 1 \pm k$ and $1-k<0$ so that

$$\mathbb{P}(|X-\mu| \leq k\sigma) = \mathbb{P}(X \leq \mu + k\sigma) = F(\mu + k\sigma) = 1 - e^{-(1+k)}$$

## 2.

$$\mathbb{P}(X \geq 2\lambda) = \mathbb{P}(X - \lambda \geq \lambda) = \mathbb{P}(|X-\lambda| \geq \lambda) \leq 1/\lambda.$$

## 3.

First, note that $\mathbb{V}(\overline{X}) = \mathbb{V}(X_1)/n = p(1-p)/n$. Chebyshev's inequality yields

$$\mathbb{P}(|\overline{X} - p| > \epsilon) \leq \frac{p\,(1-p)}{n\epsilon^2} \leq \frac{\max\{x\,(1-x) : 0 \leq x \leq 1\}}{n\epsilon^2} = \frac{1}{4n\epsilon^2}.$$

Next, note that

$$\mathbb{P}(|\overline{X} - p| \geq \epsilon) = \mathbb{P}(\overline{X} - p \geq \epsilon) + \mathbb{P}(\overline{X} - p \leq -\epsilon).$$

Let $Y_i = (X_i - \mathbb{E}X_1)/n = (X_i - p)/n$ so that $\overline{X} - p = \sum_i Y_i$. Then, $\mathbb{E}Y_i = 0$ and $-p/n \leq Y_i \leq (1-p)/n$. Hoeffding's inequality yields

$$\begin{aligned}\mathbb{P}(\overline{X} - p \geq \epsilon) = \mathbb{P}\left(\sum_i Y_i \geq \epsilon\right) &\leq \exp(-t\epsilon) \prod_i \exp\left(\frac{t^2}{8n^2}\right) \\ &= \exp\left(\frac{t^2}{8n} - t\epsilon\right) \leq \min_{t>0} \exp\left(\frac{t^2}{8n} - t\epsilon\right) = \exp(-2n\epsilon^2).\end{aligned}$$

Similarly, $\mathbb{P}(\overline{X} - p \leq -\epsilon) = \mathbb{P}(\sum_i (-Y_i) \geq \epsilon) \leq \exp(-2n\epsilon^2)$. It follows that

$$\mathbb{P}(|\overline{X} - p| \geq \epsilon) \leq 2\exp(-2n\epsilon^2) = \frac{1}{1/2 + n\epsilon^2 + n^2\epsilon^4 + O(n^4)}$$

is tighter than the Chebyshev bound for sufficiently large $n$.

## 4.

## a)

Applying our findings from Question 3,

$$\mathbb{P}(p \in C_n) = 1 - \mathbb{P}(p \notin C_n) \geq 1 - 2\exp(-2n\epsilon_n^2) = 1 - 2\exp\left(\log\left(\frac{\alpha}{2}\right)\right) = 1 - \alpha.$$

## b)

TODO (Computer Experiment)

## c)

The length of the interval is $2\epsilon_n$. This length is at most $c > 0$ if and only if $n \geq 2\log(2/\alpha)/c^2$.

TODO (Plot)

# 5.

As per the hint,

$$\begin{aligned}\mathbb{P}(|Z| > t) = 2\mathbb{P}(Z \geq t) &= \sqrt{\frac{2}{\pi}} \int_t^\infty \exp\left(-\frac{x^2}{2}\right) dx \\ &\leq \sqrt{\frac{2}{\pi}} \frac{1}{t} \int_t^\infty x \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{\frac{2}{\pi}} \frac{1}{t} \exp\left(-\frac{t^2}{2}\right).\end{aligned}$$

# 6.

TODO (Plot)

# 7.

A linear combination of IID normal random variables is itself a normal random variable. Therefore, $\overline{X}$ is a random variable with zero mean and variance $1/n$. Letting $Z \sim N(0,1)$, Mill's inequality yields

$$\mathbb{P}(|\overline{X}| \geq t) = \mathbb{P}(|Z| \geq t\sqrt{n}) \leq \sqrt{\frac{2}{\pi}} \frac{1}{t\sqrt{n}} \exp\left(-\frac{t^2 n}{2}\right).$$

The above is tighter than the Chebyshev bound $1/(t^2 n)$ for sufficiently large $n$.