**DATA 1010**
**In-class exercises**
**Samuel S. Watson**
**24 October 2018**

## Problem 1

Chebyshev's inequality says that the probability that $X$ deviates from its mean by more than $k$ standard deviations is no greater than $1/k^2$.

The U.S. mint produces dimes with an average diameter of 0.5 inches and standard deviation 0.01. Using Chebyshev's inequality, give a lower bound for the number of coins in a lot of 400 coins that are expected to have a diameter between 0.48 and 0.52.

## Solution

Let $X$ be the diameter of a dime; then $\mathbb{E}[X] = 0.5$ and $\sigma(X) = 0.01$. The interval $(0.48, 0.52)$ represents the set of numbers which are two standard deviations or less from $\mathbb{E}[X]$. Chebyshev's inequality says that

$$\mathbb{P}(|X - \mathbb{E}[X]) \geq 2\sigma) \leq \frac{1}{4}.$$

Taking the complement, we have

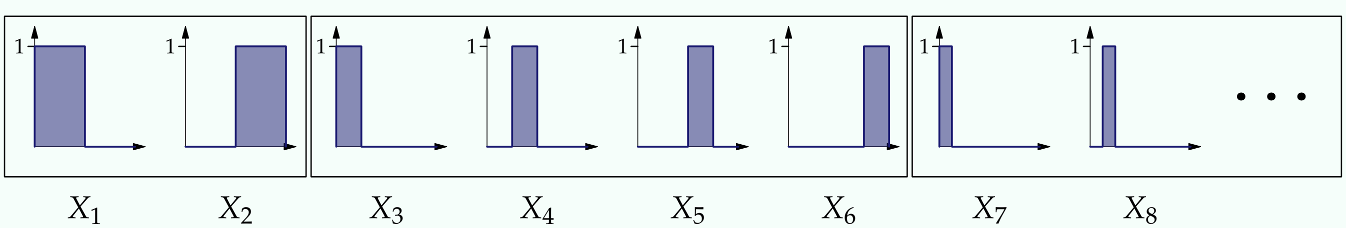
$$\begin{aligned}\mathbb{P}(|X - \mathbb{E}[X] < 2\sigma) &\geq 1 - \frac{1}{4} \\ &= \frac{3}{4}\end{aligned}$$

Because the probability of a dime having a diameter between 0.48 and 0.52 is at least $\frac{3}{4}$, the Chebyshev lower bound for the number of dimes whose diameters lie in this interval is $400 \cdot \frac{3}{4} = 300$.

## Problem 2

We say that a sequence of random variables $X_1, X_2, \ldots$ converges to a random variable $X$ in probability if $\mathbb{P}(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$, for any $\epsilon > 0$.

Consider the sequence of random variables defined on $\Omega = [0, 1]$ (with probability measure $\mathbb{P}(A) = \text{length}(A)$) as follows:




In other words, the first two random variables are the indicator functions of the first half and the last half of the unit interval. The next four are the indicators of the first quarter, the second quarter, the third quarter, and the fourth quarter of the unit interval. The next eight are indicators of width-one-eighth intervals sweeping across the interval, and so on. For concreteness, suppose that each random variable is equal to 1 at any point $\omega$ where the random variable is discontinuous.

Show that $X_n \to 0$ in probability.

**Solution**

Consider any value of $\epsilon$ between 0 and 1. The probability that $X_n$ differs from 0 by more than $\epsilon$ is $\frac{1}{2}$ for $n \in \{1, 2\}$, and $\frac{1}{4}$ for $n \in \{3, 4, 5, 6\}$, and so on. The sequence

$$\frac{1}{2}, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \ldots,$$

converges to 0, so the sequence satisfies the definition of convergence to zero in probability.

**Problem 3**

Define $f_n(x) = n\mathbf{1}_{0 \le x \le 1/n}$, and let $\nu_n$ be the probability measure with density $f_n$. Show that $\nu_n$ converges to the probability measure $\nu$ which puts of all its mass at the origin.

**Solution**

Suppose $I = (a, b)$ is a continuity interval of $\nu$ (that is, an interval with the property that $\nu$ assigns no mass to either endpoint).

If $I$ contains the origin, then the terms of sequence $\nu_1(I), \nu_2(I), \ldots$ are equal to 1 for large enough $n$, since all of the probability mass of $\nu_n$ is in the interval $\left[0, \frac{1}{n}\right]$ and eventually $\left[0, \frac{1}{n}\right] \subset I$.

If $I$ does not contain the origin, then the terms of the sequence $\nu_1(I), \nu_2(I), \ldots$ are eventually equal to 0, for the same reason.

In either case, $\nu_n(I)$ converges to $\nu(I)$. Therefore, $\nu_n$ converges to $\nu$.

**Problem 4**

Use Chebyshev's inequality to show that if $X_1, X_2, \ldots$ is a sequence of independent samples from a finite-variance distribution $\nu$, then the running average of the $X_i$'s converges to the mean of $\nu$.

**Solution**

The mean and variance of $A_n = (X_1 + \cdots + X_n)/n$ are $\mu$ and $\sigma^2/n$, where $\mu$ is the mean and $\sigma^2$ is the variance of $\nu$. Therefore, Chebyshev's inequality implies that

$$\mathbb{P}(|A_n - \mu| \ge \epsilon) \le \epsilon^{-2}\sigma^2/n.$$

Since $\epsilon^{-2}\sigma^2/n$ as $n \to \infty$, we conclude that $\mathbb{P}(|A_n - \mu| \ge \epsilon)$ converges to 0 as $n \to \infty$.

**Problem 5**

Suppose we flip a coin which has probability 60% of turning up heads $n$ times. Use the normal approximation to estimate the value of $n$ such that the proportion of heads is between 59% and 61% with probability approximately 99%.

**Solution**

We calculate the standard deviation $\sigma = \sqrt{(0.4)(0.6)}$ and the mean $\mu = 0.6$ of each flip, and we use these values to rewrite the desired probability in terms of $S_n^*$. We find

$$\begin{aligned} P\left(0.59 < \frac{1}{n}S_n < 0.61\right) &= P\left(-0.01 < \frac{S_n - \mu n}{n} < 0.01\right) \\ &= P\left(-\frac{0.01\sqrt{n}}{\sqrt{0.4 \cdot 0.6}} < \frac{S_n - \mu n}{\sigma\sqrt{n}} < \frac{0.01\sqrt{n}}{\sqrt{0.4 \cdot 0.6}}\right), \end{aligned}$$

where the last step was obtained by multiplying all three expressions in the compound inequality by $\sqrt{n}/\sigma$. Since $S_n^*$ is distributed approximately like a standard normal random variable, the normal approximation tells us to look for the least $n$ so that

$$\int_{-a_n}^{a_n} dt > 0.99, \tag{1}$$

where $a_n = 0.01\sqrt{n}/\sqrt{0.4 \cdot 0.6}$. By the symmetry of the Gaussian density, we may rewrite (1) as

$$\int_{-\infty}^{a_n} dt > 0.995.$$

Defining the normal CDF $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$, we want to find the least integer $n$ such that $a_n$ exceeds $\Phi^{-1}(0.995)$. The following code tells us that $\Phi^{-1}(0.995) \approx 2.576$.

```
using Distributions
quantile(Normal(0,1),0.995)
```

Setting this equal to $a_n$ and solving for $n$ gives 15,924. This approximation is quite close to the exact value of 15,861.