Recitation 21 Answers
May 3, 2005

1. Note that n is deterministic and H is a random variable.

   (a) Use X_1, X_2, ... to denote the (random) measured heights.

$$\begin{aligned} H &= \frac{X_1 + X_2 + \cdots + X_n}{n} \\ \mathbf{E}[H] &= \frac{\mathbf{E}[X_1 + X_2 + \cdots + X_n]}{n} = \frac{n\mathbf{E}[X]}{n} = h \\ \sigma_H &= \sqrt{\mathrm{var}(H)} = \sqrt{\frac{n\,\mathrm{var}(X)}{n^2}} \quad \text{(var of sum of \textit{independent} r.v.s is sum of vars)} \\ &= \frac{1.5}{\sqrt{n}} \end{aligned}$$

   (b) We solve $\frac{1.5}{\sqrt{n}} < 0.01$ for n to obtain $n > 22500$.

   (c) Apply the Chebyshev inequality to H with $\mathbf{E}[H]$ and $\mathrm{var}(H)$ from part (a):

$$\begin{aligned} \mathbf{P}(|H - h| \geq t) &\leq \left(\frac{\sigma_H}{t}\right)^2 \\ \mathbf{P}(|H - h| < t) &\geq 1 - \left(\frac{\sigma_H}{t}\right)^2 \end{aligned}$$

   To be "99% sure" we require the latter probability to be at least 0.99. Thus we solve

$$1 - \left(\frac{\sigma_H}{t}\right)^2 \geq 0.99$$

   with $t = 0.05$ and $\sigma_H = \frac{1.5}{\sqrt{n}}$ to obtain

$$n \geq \left(\frac{1.5}{0.05}\right)^2 \frac{1}{0.01} = 90000.$$

   (d) The variance of a random variable increases as its distribution becomes more spread out. In particular, if a random variable is known to be limited to a particular closed interval, the variance is maximized by having 0.5 probability of taking on each endpoint value. In this problem, random variable X has an unknown distribution over $[0, 3]$. The variance of X cannot be more than the variance of a random variable that equals 0 with probability 0.5 and 3 with probability 0.5. This translates to the standard deviation not exceeding 1.5.
   In fact, this argument can be made more rigorous as follows.
   First, we have

$$\mathrm{var}(X) \leq \mathbf{E}[(X - \frac{3}{2})^2] = \mathbf{E}[X^2] - 3\mathbf{E}[X] + \frac{9}{4}$$

   since $\mathbf{E}[(X - a)^2]$ is minimized when a is the mean (i.e., the mean is the least-squared estimator).
   Second, we also have

$$0 \leq \mathbf{E}[X(3 - X)] = \mathbf{E}[X] - \mathbf{E}[X^2]$$

since the variable has support in $[0, 3]$. Adding the above two inequalities, we have

$$\text{var}(X) \leq \frac{9}{4}$$

or equivalently, $\sigma_X \leq \frac{3}{2}$.

2. (a) We want to find the probability that there are at least 45 successes out of 50 total trials, where the probability of success is given to be .95. Using the Normal approximation to the binomial (where $\mu = 47.5$ and $\sigma \approx 1.54$), we find:

$$\begin{aligned}
\mathbf{P}(45 \text{ to } 50 \text{ successes}) &\approx 1 - \Phi\left(\frac{44.5 - \mu}{\sigma}\right) \\
&\approx 1 - \Phi(-1.95) \\
&= \Phi(1.95) \\
&= 0.9744
\end{aligned}$$

(b) To be able to use the Poisson approximation p has to be small and n has to be relatively large. Therefore, using $p = 0.95$ will not give a good approximation. Instead, we define a new random variable, I, to be the number of incorrect predictions out of 50.

$$\begin{aligned}
\mathbf{P}(45 \text{ to } 50 \text{ successes}) &= \mathbf{P}(I = 0) + \mathbf{P}(I = 1) + \ldots + \mathbf{P}(I = 5) \\
&\approx \sum_{k=0}^{5} \frac{2.5^k e^{-2.5}}{k!} \approx 0.9582
\end{aligned}$$

The second method, although more tedious, is perhaps more appropriate. The Normal approximation works well with sums of symmetric distributions, which for the binomial is satisfied when p is close to .5. Here p is quite far from that. Of course, the Normal distribution makes it quite convenient to calculate, especially when the number of terms in the sum grows.

3. (a) $\mathbf{E}[X_n] = 0 * (1 - \frac{1}{n}) + 1 * \frac{1}{n} = \frac{1}{n}$
$\text{var}(X_n) = (0 - 1/n)^2 * (1 - 1/n) + (1 - 1/n)^2 * (1/n) = \frac{n-1}{n^2}$
$\mathbf{E}[Y_n] = 0 * (1 - \frac{1}{n}) + n * \frac{1}{n} = 1$
$\text{var}(Y_n) = (0 - 1)^2 * (1 - 1/n) + (n - 1)^2 * (1/n) = n - 1$

(b) Using the Chebyshev inequality, we have
$\lim_{n->\infty} \mathbf{P}(|X_n - \frac{1}{n}| \geq \epsilon) \leq \lim_{n->\infty} \frac{n-1}{n^2\epsilon^2} = 0$
Moreover, $\lim_{n->\infty} \frac{1}{n} = 0$.
it follows that X_n is convergent in probability.

(c) In this case, Chebyshev suggests that
$\lim_{n->\infty} \mathbf{P}(|Y_n - 1| \geq \epsilon) \leq \lim_{n->\infty} \frac{n-1}{\epsilon^2} = \infty$,
It follows that the Chebyshev inequality tells us nothing.

(d) For every $\epsilon > 0$, we have
$\lim_{n->\infty} \mathbf{P}(|X_n| \geq \epsilon) \leq \lim_{n->\infty} \frac{1}{n} = 0$,
which suggests that X_n converges to zero in probability.

(e) For every $\epsilon > 0$, we have
$\lim_{n->\infty} \mathbf{P}(|Y_n| \geq \epsilon) \leq \lim_{n->\infty} \frac{1}{n} = 0$,
which suggests that Y_n converges to zero in probability.