Homework 2

Daniel Hartig

September 13, 2017

Problem 1.2

a.
Each randomly guessed question is a Bernoulli trial with success probability 0.25. Since each trial is independent, the distribution of the number of correct answers is Binomial with parameters $\pi = 0.25$ and $n = 100$.

b.
The mean of a binomial distribution is $n\pi = 100(0.25) = 25$. The standard deviation is $\sqrt{n\pi(1-\pi)} = 4.330$. If the student picked 50 answers correctly, that would be $\frac{50-25}{4.330} = 5.774$ standard deviations above the mean; a very unusual event. Calculating probabilities from the binomial distribution, there is a $\binom{100}{50}0.25^{50}(1-0.25)^{50} = 4.507 \times 10^{-8}$ chance of the student selecting exactly 50 answers correctly, and a 6.639×10^{-8} change of selecting 50 or more choices correctly.

c.
Since the students choice of answer (1, 2, 3, or 4) is random, $\sum_{j=1}^{4} \pi_i = 1$, and each choice is selected with equal probability ($\pi_1 = \pi_2 = \pi_3 = \pi_4$). Therefore, $\pi_1 = \pi_2 = \pi_3 = \pi_4 = 1/4$, and n_j is a four dimensional multinomial distribution where $\pi_i = 0.25; \forall i \in \{1, 2, 3, 4\}$. Alternately, n_j could be described as a discrete uniform distribution with parameters $a = 1, b = 4$.

d.
For all $j \in \{1, 2, 3, 4\}$,

$$E[n_j] = n\pi_j = 25$$
$$\mathrm{Var}(n_j) = n\pi_j(1-\pi_j) = 18.75$$

For all $j, k \in \{1, 2, 3, 4\}; j \neq k$,

$$\mathrm{Cov}(n_j, n_k) = -n\pi_j\pi_k = -6.25$$
$$\mathrm{Corr}(n_j, n_k) = \frac{\mathrm{Cov}(n_j, n_k)}{\sqrt{\mathrm{Var}(n_j)}\sqrt{\mathrm{Var}(n_k)}} = \frac{-6.25}{\sqrt{18.75}\sqrt{18.75}} = -1/3$$

Problem 1.16

The variance of a binomial distribution is $n\pi(1-\pi) = -\pi^2 - \pi$, which is the equation for an inverted parabola with a maximum at $\pi = 0.5$ and minimum– on the range $(0, 1)$–at $\pi = 0$ and $\pi = 1$. Thus, the variance of a binomial distribution is highest when $\pi = 0.5$ and lowest near $\pi = 0$ and $\pi = 1$.

The width of the confidence interval for any estimate of a parameter is proportional to the variance of that parameter and inversely proportional to the number of samples. Thus, if π is close to 0 or 1, fewer samples will be needed to get yield an acceptably small confidence interval.

Problem 1.17

a.
There are two possible outcomes for Y_i: 0 and 1. $P(Y_i = 0) + P(Y_i = 1) = 1$. Therefore, Y_i is a Bernoulli trial with success defined as $Y_i = 1$ and probability of success $P(Y_i = 1) = \pi$. Y is the total number of successes in n Bernoulli trials, so the distribution of Y is binomial. For a binomial distribution with paramenters π and n, $E[Y] = n\pi$ and $\mathrm{Var}(Y) = n\pi(1-\pi)$.

b.

$$\begin{aligned}\mathrm{Var}(Y) &= \mathrm{Var}\left(\sum_{i=1}^{n} Y_i\right)\\ &= \sum_{i\in 1,..,n} \mathrm{Var}(Y_i) + 2\sum_{i\in 1,..,n}\sum_{\substack{j\in 1,..,n\\ i\neq j}} \mathrm{Cov}(Y_i, Y_j)\\ &= n\pi(1-\pi) + 2\sum_{i\in 1,..,n}\sum_{\substack{j\in 1,..,n\\ i\neq j}} \mathrm{Cov}(Y_i, Y_j).\end{aligned}$$

Since any pair Y_i, Y_j has covariance $\rho > 0$,

$$2\sum_{i\in 1,..,n}\sum_{\substack{j\in 1,..,n\\ i\neq j}} \mathrm{Cov}(Y_i, Y_j) > 0$$

and

$$n\pi(1-\pi) + 2\sum_{i\in 1,..,n}\sum_{\substack{j\in 1,..,n\\ i\neq j}} \mathrm{Cov}(Y_i, Y_j) > n\pi(1-\pi).$$

c.

$$\begin{aligned}\mathrm{Var}(Y_i) &= E\left[\mathrm{Var}(Y_i|\pi)\right] + \mathrm{Var}\left(E[Y_i|\pi]\right)\\ &= E\left[\pi(1-\pi)\right] + \mathrm{Var}(\pi)\\ &= \rho(1-\rho) + \mathrm{Var}(\pi)\end{aligned}$$

Since $\mathrm{Var}(\pi) > 0$,

$$\begin{aligned}\mathrm{Var}(Y) &= n\mathrm{Var}(Y_i)\\ &= n\rho(1-\rho) + n\mathrm{Var}(\pi) > n\rho(1-\rho)\end{aligned}$$