Due Date: October 17th, 2019 *Problems Due: 1, 3(b), 3(c), 4(a)*

1. Suppose there 5000 balls in an urn of which 1000 are pink and 4000 are green.
   (a) Natasha chooses 4 balls *with replacement*. Using `dbinom` command in R find the probability that there are exactly 2 of the balls chosen are pink.
   (b) Salima chooses 4 balls *without replacement*. Using `dhyper` command in R find the probability that there are exactly 2 of the balls chosen are pink.
2. Let N, m, and r be positive integers for which $m < r < N$ and let k be a positive integer between 0 and m. Define
$$p = \frac{r}{N}, \quad p_1 = \frac{r-k}{N-k}, \quad \text{and} \quad p_2 = \frac{r-k}{N-m}.$$
   (a) Show that
$$\binom{m}{k} p_1^k (1-p_2)^{m-k} < \frac{\binom{r}{k}\binom{N-r}{m-k}}{\binom{N}{m}} \leq \binom{m}{k} p^k (1-p_1)^{m-k}.$$
   (b) Can you now justify the statement : – *when samples are small relative to the size of the populations they came from, the two methods of sampling (with and without replacement) give very similar results.*
3. Suppose there N balls in an urn of which B are pink and rest are green. Let $p = \frac{B}{N}$ and $P = 100p$. Suppose we choose m balls *with* replacement. Let $X_1, X_2, \ldots, X_m$ be random variables that take value 1 if a pink ball is chosen and 0 otherwise respectively in m draws. Let $\hat{p}_m = \bar{X}$ and $\hat{P}_m = 100\hat{p}_m$.
   (a) Show that for
$$\mathbb{P}(\mid P - \hat{P}_m \mid < \alpha\%) \geq 1 - \frac{10000}{4m\alpha^2}$$
   (b) Using R or otherwise calculate m for which the probability that, $\mid P - \hat{P}_m \mid$ is within $\alpha = 1, 2, 5\%$, is .99.
   (c) Assume $P = 60$, using `pbinom` calculate for α and (respective) m as above calculate $\mathbb{P}(\mid P - \hat{P}_m \mid < \alpha\%)$. Compare with the answer from the bounds in previous part.
4. Suppose there N balls in an urn of which B are pink and rest are green. Let $p = \frac{B}{N}$ and $P = 100p$. Suppose we choose m balls *without* replacement. Let $\hat{p}_m$ be the proportion of pink balls in the sample and $\hat{P}_m = 100\hat{p}_m$.
   (a) Using R simulate 100 trials of the above experiment when $N \in \{100000, 400000, 1000000, 4000000\}$ and for $m \in \{1000, 2500, 4000, 6250, 8000\}$ and compute the number of trials in which $\mid P - \hat{P}_m \mid < 2\%$.
   (b) *Extra Credit:* Using the Python program, provided by Rajeeva Karandikar, in the shared dropbox folder compute the exact $\mathbb{P}(\mid P - \hat{P}_m \mid < 2\%)$ for N and m in the previous part.
5. (*Pareto Distribution*) Let $\beta > 1, \alpha > 0$ and $X_1, X_2, \ldots, X_m$ be i.i.d. random variables with commmon probability density function
$$f(x \mid \alpha, \beta) = \begin{cases} \beta\alpha^\beta x^{-(\beta+1)} & \text{if } x > \alpha \\ 0 & \text{otherwise} \end{cases}$$
   (a) Assume $\alpha = 1$, find the $P(X_1 > x)$ for $x \in \mathbb{R}$.
   (b) Assume $\beta > 1$ is a known constant. Find the method of moments estimator for α.
   (c) Assume $\beta > 2$. Can you provide the method of moments estimators for α, β ?
6. (*Weibull Distribution*) Let $0 < \alpha, \beta$ and $X_1, X_2, \ldots, X_m$ be i.i.d. random variables with commmon probability density function
$$f(x \mid \alpha, \beta) = \begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^\beta} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$
   Assume $\beta > 0$ is a known constant. Find the maximum likelihood estimator for α.