1. Show that if X is a continuous random variable, then

$$\min_a \mathbb{E}\,|X-a| = \mathbb{E}\,|X-m|,$$

where m is the median of X.

Let $g(a) = \mathbb{E}\,|X-a|$ be a function of a. We now find the minimum of a. Suppose X has a cumulative distribution $F(\cdot)$, then

$$\begin{aligned} g(a) = \mathbb{E}\,|X-a| &= \int_{-\infty}^{\infty} |x-a| dF(x) \\ &= \int_{-\infty}^{a} (a-x)dF(x) + \int_{a}^{\infty} (x-a)dF(x) \\ &= a(F(a)-0) - \int_{-\infty}^{a} xdF(x) + \int_{a}^{\infty} xdF(x) - a(1-F(a)) \\ &= 2aF(a) - a - \int_{-\infty}^{a} xdF(x) + \int_{a}^{\infty} xdF(x). \end{aligned}$$

Using the fundamental theorem of calculus, we have

$$g'(a) = 2F(a) + 2aF'(a) - 1 - aF'(a) - aF'(a) = 2F(a) - 1.$$

Let $g'(m) = 0$ we have $F(m) = 1/2$ which means that m, the minimizer of function g, is the median of X.

2. Let X and Y be independent standard normal random variables.

   (a) Show that $\frac{X}{X+Y}$ has a Cauchy distribution.

The joint distribution of X and Y is

$$f_{X,Y}(x,y) = \frac{1}{2\pi} \exp\left(-\frac{x^2}{2}\right) \exp\left(-\frac{y^2}{2}\right).$$

Let $M_1 = X/(X+Y) \in \mathbb{R}$ and $M_2 = X+Y \in \mathbb{R}$, then we have $X = M_1M_2$ and $Y = M_1M_2 - M_2$, so the Jacobian matrix of this transformation is

$$\begin{bmatrix} M_2 & M_1 \\ M_2 & M_1 - 1 \end{bmatrix}$$

with determinant $-M_2$, hence by transformation of random variables formula,

$$\begin{aligned} f_{M_1,M_2}(m_1,m_2) &= \frac{1}{2\pi} \exp\left(-\frac{m_1^2m_2^2}{2}\right) \exp\left(-\frac{m_1^2m_2^2 - 2m_1m_2^2 + m_2^2}{2}\right) m_2 \\ &= \frac{1}{2\pi} m_2 \exp\left(-\frac{2m_1^2 - 2m_1 + 1}{2} m_2^2\right). \end{aligned}$$

The marginal density of M_1 is therefore

$$\begin{aligned} f_{M_1}(m_1) &= \int_{-\infty}^{\infty} f_{M_1,M_2}(m_1,m_2) dm_2 \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{2m_1^2 - 2m_1 + 1}{2} m_2^2\right) dm_2^2 \\ &= \int_{0}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{2m_1^2 - 2m_1 + 1}{2} x\right) dx \\ &= \frac{1}{\pi(2m_1^2 - 2m_1 + 1)}. \end{aligned}$$

Hence M_1 follows a Cauchy distribtuion with location $1/2$ and scale $1/2$.

   (b) Find the distribution of $X/|Y|$.

We first show that the distribution of $X/|Y|$ is identical to the distribution of X/Y. Note that

$$\Pr(X/|Y| \le z) = \Pr(X \le z|Y|, Y \ge 0) + \Pr(X \le z|Y|, Y < 0)$$

$$
\begin{aligned}
&= \Pr(X \le zY, Y \ge 0) + \Pr(X \le -zY, Y < 0) \\
\text{(by the symmetric of } X) &= \Pr(X \le zY, Y \ge 0) + \underline{\Pr(-X \le -zY, Y < 0)} \\
&= \Pr(X \le zY, Y \ge 0) + \Pr(X \ge zY, Y < 0) \\
&= \Pr(X/Y \le z, Y \ge 0) + \Pr(X/Y \le z, Y < 0) \\
&= \Pr(X/Y \le z).
\end{aligned}
$$

Let $M_1 = X/Y \in \mathbb{R}$, $M_2 = Y \in \mathbb{R}$, then $X = M_1M_2$, $Y = M_2$, so the Jacobian matrix of this transformation is

$$
\begin{bmatrix} M_2 & M_1 \\ 0 & 1 \end{bmatrix}
$$

with determinant M_2, hence by transformation of random variables formula,

$$
\begin{aligned}
f_{M_1,M_2}(m_1, m_2) &= \frac{1}{2\pi} \exp\left(-\frac{m_1^2 m_2^2}{2}\right) \exp\left(-\frac{m_2^2}{2}\right) m_2 \\
&= \frac{1}{2\pi} m_2 \exp\left(-\frac{m_1^2+1}{2} m_2^2\right).
\end{aligned}
$$

The marginal density of M_1 is therefore

$$
\begin{aligned}
f_{M_1}(m_1) &= \int_{-\infty}^{\infty} f_{M_1,M_2}(m_1, m_2) dm_2 \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{m_1^2+1}{2} m_2^2\right) dm_2^2 \\
&= \int_{0}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{m_1^2+1}{2} x\right) dx \\
&= \frac{1}{\pi(m_1^2+1)}.
\end{aligned}
$$

Hence M_1 follows a Cauchy distribtuion with location 0 and scale 1.

3. Axiom of countable additivity: For any countably infinite sequence of events $(A_i)_{i\ge1}$ with $A_i \cap A_j = \emptyset$ when $i \ne j$, then $\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$.

   Finite additivity: For any finite sequence of events $(A_i)_{i=1,2,\ldots,n}$ with $A_i \cap A_j = \emptyset$ when $i \ne j$, then $\Pr(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \Pr(A_i)$.

   (a) Show that the Axiom of Countable Additivity implies Finite Additivity.

Suppose we have a countably infinite sequence of events $(A_i)_{i\ge1}$ with $A_i \cap A_j = \emptyset$ when $i \ne j \in \{1, 2, \ldots, n\}$, and $A_k = \emptyset$ for $k = n+1, n+2, \ldots$, then this sequence satisfy the condition that $A_i \cap A_j = \emptyset$ when $i \ne j$, so we have

$$
\begin{aligned}
\Pr(\bigcup_{i=1}^{n} A_i) &= \Pr(\bigcup_{i=1}^{n} A_i \cup A_{n+1} \cup A_{n+2} \cup \ldots) \\
&= \Pr(\bigcup_{i=1}^{\infty} A_i) \\
\text{(by countable additivity)} &= \sum_{i=1}^{\infty} \Pr(A_i) \\
&= \sum_{i=1}^{n} \Pr(A_i) + \sum_{i=n+1}^{\infty} \Pr(A_i) \\
&= \sum_{i=1}^{n} \Pr(A_i).
\end{aligned}
$$

   (b) Let $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$ be an infinite sequence of nested sets whose limit is the empty set, which we denote by $A_n \downarrow \emptyset$. Axiom of Continuity means if $A_n \downarrow \emptyset$, then $P(A_n) \to 0$. Prove that the Axiom of Continuity and the Axiom of Finite Additivity together imply Countable Additivity.

Suppose we have a countably infinite sequence of events $(A_i)_{i\geq 1}$ with $A_i \cap A_j = \emptyset$ when $i \neq j$. Let $B_n = \cup_{i=n+1}^{\infty} A_i$, then we know that $B_n \supset B_{n+1} \supset \dots$ with $B_n \downarrow \emptyset$. By axiom of continuity, we have $\lim_{n\to\infty} \Pr(B_n) = 0$. Then we have

$$\begin{aligned}
\Pr(\bigcup_{i=1}^{\infty} A_i) &= \lim_{n\to\infty} \Pr(\bigcup_{i=1}^{n} A_i \cup B_n) \\
&= \lim_{n\to\infty} \Pr(\bigcup_{i=1}^{n} A_i) + \lim_{n\to\infty} \Pr(B_n) \\
\text{(by finite additivity)} &= \lim_{n\to\infty} \sum_{i=1}^{n} \Pr(A_i) + 0 \\
&= \sum_{i=1}^{\infty} \Pr(A_i).
\end{aligned}$$

4. Prove

$$\lim_{n\to\infty} \frac{n!}{n^{(n+1/2)}e^{-n}} = C,$$

where C is a positive constant. You shouldn't use Stirling's Formula to prove the claim.

Let

$$a_n = \frac{n!}{n^{(n+1/2)}e^{-n}}, \qquad b_n = \log a_n = \sum_{i=1}^{n} \log i - \left(n + \frac{1}{2}\right) \log n + n.$$

We show that b_n is monotonely decreasing and bounded below by a positive constant and so is a_n, so by elementary mathematical analysis, a_n converges to a limit $C > 0$. Without finding the exact value of C since it is not required we give the following proof. Note that

$$\begin{aligned}
b_n - b_{n+1} &= \left(\sum_{i=1}^{n} \log i - \left(n + \frac{1}{2}\right) \log n + n\right) - \left(\sum_{i=1}^{n+1} \log i - \left(n + 1 + \frac{1}{2}\right) \log(n+1) + n + 1\right) \\
&= \left(n + 1 + \frac{1}{2}\right) \log(n+1) - \left(n + \frac{1}{2}\right) \log n - \log(n+1) - 1 \\
&= \left(n + \frac{1}{2}\right) \log\left(\frac{n+1}{n}\right) - 1.
\end{aligned}$$

Consider the Taylor expansion for $\log((1+x)/(1-x))$ for $|x| < 1$:

$$\log\left(\frac{1+x}{1-x}\right) = 2\sum_{i=1}^{\infty} \frac{x^{2i-1}}{2i-1},$$

and let $x = 1/(2n+1) < 1$, we have

$$\log\left(\frac{n+1}{n}\right) = 2\sum_{i=1}^{\infty} \frac{1}{(2i-1)(2n+1)^{2i-1}},$$

so

$$\begin{aligned}
b_n - b_{n+1} &= \left(n + \frac{1}{2}\right) \cdot 2\sum_{i=1}^{\infty} \frac{1}{(2i-1)(2n+1)^{2i-1}} - 1 \\
&= \sum_{i=1}^{\infty} \frac{1}{(2i-1)(2n+1)^{2i-2}} - 1 \\
&= \sum_{i=1}^{\infty} \frac{1}{(2i+1)(2n+1)^{2i}} > 0,
\end{aligned}$$

so (b_n) is a decreasing sequence.

Note also that

$$b_n - b_{n+1} = \sum_{i=1}^{\infty} \frac{1}{(2i+1)(2n+1)^{2i}}$$

$$
\begin{aligned}
&< \frac{1}{3}\sum_{i=1}^{\infty}\frac{1}{(2n+1)^{2i}} \\
&= \frac{1}{3}\frac{\frac{1}{(2n+1)^2}}{1-\frac{1}{(2n+1)^2}} \\
&= \frac{1}{3}\frac{1}{(2n+1)^2-1} \\
&= \frac{1}{12}\frac{1}{n(n+1)} \\
&= \frac{1}{12}\left(\frac{1}{n}-\frac{1}{n+1}\right),
\end{aligned}
$$

so $(b_n - 1/12n)$ is a increasing sequence with lower bound $b_1 - 1/12 = 11/12 > 0$, which means $b_n > 11/12$. Combining the fact that (b_n) is decreasing and bounded below, we know that (b_n) is converging to a positive constant c, and so is (a_n) to $C = \exp(c)$.

5. Prove, by definition, that if $A_n \xrightarrow{p} 1$ and $Y_n \xrightarrow{d} Y$, then $A_nY_n \xrightarrow{d} Y$

Note that $A_nY_n = Y_n(A_n-1) + Y_n$. It suffices to show $Y_n(A_n-1) \xrightarrow{p} 0$. Then we obtain the convergence by the additive version. Since $Y_n \xrightarrow{d} Y$, $Y_n = O_p(1)$. For $\forall\ \delta,\ \epsilon > 0$, we could find M s.t. $\sup \mathbf{P}(|Y_n| \geq M) \leq \frac{\epsilon}{2}$, and N s.t. $\mathbf{P}(|A_n - 1| \geq \frac{\delta}{M}) \leq \frac{\epsilon}{2}$ hold for $\forall\ n > N$. Then we have $\forall\ n > N$,

$$
\begin{aligned}
\mathbf{P}(|Y_n(A_n-1)| \geq \delta) \leq& \mathbf{P}(|Y_n| \geq M) + \mathbf{P}(|A_n - 1| \geq \frac{\delta}{M}) \\
\leq& \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon
\end{aligned}
$$

Therefore, $Y_n(A_n-1) \xrightarrow{p} 0$.

Remark:

- $Y_n = O_p(1)$ because we could find the corresponding uniform constant M with asymptotic properties. When n is large enough, Y_n is close to Y. Then find a sufficiently large M for the first finite ones.

6. Let $X_1, \ldots, X_n$ be independently distributed with exponential density

$$\frac{1}{2\theta}\exp(-x/2\theta)I(x \geq 0),$$

and let the ordered X's be denoted by $Y_1 \leq Y_2 \leq \cdots \leq Y_n$. It is assumed that Y_1 becomes available first, then Y_2, and so on, and that observation is continued until Y_r has been observed. This might arise, for example, in life testing where each X measures the length of life of, say, an electron tube, and n tubes are being tested simultaneously. Another application is to the disintegration of radioactive material, where n is the number of atoms, and observation is continued until r α-particles have been emitted.

(a) Show that the joint distribution of $Y_1 \leq Y_2 \leq \cdots \leq Y_r$ has density

$$\frac{1}{(2\theta)^r}\frac{n!}{(n-r)!}\exp\left\{-\frac{\sum_{i=1}^r y_i + (n-r)y_r}{2\theta}\right\}, \qquad 0 \leq y_1 \leq y_2 \leq \cdots \leq y_r.$$

We use the formula for all n order statistics with density function $f(\cdot)$:

$$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = n! f(y_1)f(y_2)\cdots f(y_n)I(0 \leq y_1 \leq \cdots \leq y_n).$$

For the specified exponential density, it becomes

$$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = n!\frac{1}{(2\theta)^n}\exp\left(-\frac{\sum_{i=1}^n y_i}{2\theta}\right)I(0 \leq y_1 \leq \cdots \leq y_n).$$

Integrate the above joint density with respect to $y_n, y_{n-1}, \ldots, y_{r+1}$:

$$f_{Y_1,Y_2,\ldots,Y_r}(y_1, y_2, \ldots, y_r) = n!\frac{1}{(2\theta)^n}\int_{y_r}^{\infty}\cdots\int_{y_{n-2}}^{\infty}\int_{y_{n-1}}^{\infty}\exp\left(-\frac{\sum_{i=1}^n y_i}{2\theta}\right)dy_n dy_{n-1}\cdots dy_{r+1}$$

$$
\begin{aligned}
&= n!\frac{1}{(2\theta)^{n-1}}\int_{y_r}^{\infty}\cdots\int_{y_{n-2}}^{\infty}\exp\left(-\frac{\sum_{i=1}^{n-2}y_i+2y_{n-1}}{2\theta}\right)dy_{n-1}\cdots dy_{r+1}\\
&= n!\frac{1}{(2\theta)^{n-2}}\int_{y_r}^{\infty}\cdots\int_{y_{n-3}}^{\infty}\frac{1}{2!}\exp\left(-\frac{\sum_{i=1}^{n-3}y_i+3y_{n-2}}{2\theta}\right)dy_{n-2}\cdots dy_{r+1}\\
&= \dots\\
&= n!\frac{1}{(2\theta)^{r+1}}\int_{y_r}^{\infty}\frac{1}{(n-r-1)!}\exp\left(-\frac{\sum_{i=1}^{r}y_i+(n-r)y_{r+1}}{2\theta}\right)dy_{r+1}\\
&= \frac{1}{(2\theta)^r}\frac{n!}{(n-r)!}\exp\left(-\frac{\sum_{i=1}^{r}y_i+(n-r)y_r}{2\theta}\right),
\end{aligned}
$$

for $0 \le y_1 \le y_2 \le \cdots \le y_r$.

(b) Argue that the distribution of $[\sum_{i=1}^r Y_i + (n-r)Y_r]/\theta$ is χ^2 with $2r$ degrees of freedom.

We prove that the spacings $M_1 = Y_1 \in \mathbb{R}^+, M_2 = Y_2 - Y_1 \in \mathbb{R}^+, \dots, M_r = Y_r - Y_{r-1} \in \mathbb{R}^+$ are independent. From the above transformations we have $Y_1 = M_1, Y_2 = M_1 + M_2, Y_r = M_1 + \cdots + M_r$, so the Jacobian of this transformation is

$$
\begin{bmatrix}
1 & 0 & 0 & \cdots & 0\\
1 & 1 & 0 & \cdots & 0\\
1 & 1 & 1 & \cdots & 0\\
\vdots & \vdots & \vdots & \ddots & \vdots\\
1 & 1 & 1 & \cdots & 1
\end{bmatrix},
$$

with determinant 1, so the joint density of $M_1, \dots, M_r$ is

$$
\begin{aligned}
&f_{M_1,M_2,\dots,M_r}(m_1,m_2,\dots,m_r)\\
=&f_{Y_1,Y_2,\dots,Y_r}(m_1, m_1+m_2,\dots,m_1+\cdots+m_r)\\
=&\frac{1}{(2\theta)^r}\frac{n!}{(n-r)!}\exp\left(-\frac{nm_1+(n-1)m_2+\cdots+(n-r+1)m_r}{2\theta}\right)\\
=&\left[\frac{n}{2\theta}\exp\left(-\frac{nm_1}{2\theta}\right)\right]\left[\frac{n-1}{2\theta}\exp\left(-\frac{(n-1)m_2}{2\theta}\right)\right]\cdots\left[\frac{n-r+1}{2\theta}\exp\left(-\frac{(n-r+1)m_r}{2\theta}\right)\right],
\end{aligned}
$$

so M_i's are independent and each M_i has an exponential distribution with mean $2\theta/(n-i+1)$, for $i = 1, 2, \dots, r$.

Recall that if X follows an exponential distribution with mean λ, then X/λ follows an exponential distribution with unit mean, so

$$\frac{nM_1}{2\theta}, \frac{(n-1)M_2}{2\theta}, \dots, \frac{(n-r+1)M_r}{2\theta} \overset{iid}{\sim} \mathrm{Exp}(1).$$

Recall also that an exponential distribution with mean 2 is also a chi-square distribution with degree of freedom 2, we know that

$$\frac{nM_1}{\theta}, \frac{(n-1)M_2}{\theta}, \dots, \frac{(n-r+1)M_r}{\theta} \overset{iid}{\sim} \chi_2^2.$$

Surprisingly,

$$\frac{nM_1}{\theta} + \frac{(n-1)M_2}{\theta} + \cdots + \frac{(n-r+1)M_r}{\theta} = \frac{\sum_{i=1}^r Y_i + (n-r)Y_r}{\theta},$$

so we conclude that $[\sum_{i=1}^r Y_i + (n-r)Y_r]/\theta$ is χ^2 with $2r$ degrees of freedom using the property of iid chi-square random variables.

(c) Let $Y_1, Y_2, \dots$ denote the time required until the first, second, ... event occurs in a Poisson process with parameter $1/2\theta'$. Prove that $Z_1 = Y_1/\theta'$, $Z_2 = (Y_2 - Y_1)/\theta'$, $Z_3 = (Y_3 - Y_2)/\theta'$, ... are independent distributed as χ^2 with 2 degrees of freedom, and the joint density of $Y_1, \dots, Y_r$ has the density

$$\frac{1}{(2\theta')^r}\exp\left(-\frac{y_r}{2\theta'}\right), \qquad 0 \le y_1 \le y_2 \le \cdots \le y_r.$$

The distribution of Y_r/θ' is again χ^2 with $2r$ degrees of freedom.

First consider Z_1. By the property of Poisson process $N(t)$ with parameter $\lambda = 1/2\theta'$, we know that

$$\begin{aligned}\Pr(Z_1 \geq z) &= \Pr(Y_1 \geq z\theta') \\ &= \Pr(N(z\theta') = 0) \\ &= \frac{(\lambda z\theta')^0}{0!}e^{-\lambda z\theta'} \\ &= e^{-z/2},\end{aligned}$$

so $Z_1 \sim \text{Exp}(1/2)$ which is identical to χ_2^2.
Second consider Z_2. We know that

$$\begin{aligned}\Pr(Z_2 \geq z_2, Z_1 \geq z_1) &= \int_{z_1}^{\infty} \Pr(Z_2 \geq z_2 | Z_1 = z_0)\Pr(Z_1 = z_0)dz_0 \\ &= \int_{z_1}^{\infty} \Pr(Y_2 - Y_1 \geq z_2\theta' | Y_1 = z_0\theta') f_{Z_1}(z_0)dz_0 \\ &= \Pr(N(z_2\theta') = 0)\int_{z_1}^{\infty} f_{Z_1}(z_0)dz_0 \\ &= e^{-z_2/2}e^{-z_1/2},\end{aligned}$$

so $f_{Z_1,Z_2}(z_1, z_2) = (e^{-z_2/2}/2)(e^{-z_1/2}/2)$ which shows that Z_1 and Z_2 are independent, with $Z_2 \sim \text{Exp}(1/2)$ which is identical to χ_2^2. By similar arguments, we can show that $Z_3, \ldots, Z_r$ are all $\text{Exp}(1/2)$ and hence χ_2^2 distributed.
Third, since $Z_1, \ldots, Z_r$ are χ_2^2 distributed, we know that joint density of them is

$$\frac{1}{2^r}\exp\left(-\frac{\sum_{i=1}^r z_i}{2}\right).$$

From the transformation we know that $Y_1 = \theta' Z_1, Y_2 = \theta'(Z_1 + Z_2), \ldots, Y_r = \theta'(Z_1 + \ldots + Z_r)$ with $0 \leq Y_1 \leq \ldots \leq Y_r$, and the Jacobian of the transformation from $(Z_i)_{i=1,2,\ldots,r}$ to $(Y_i)_{i=1,2,\ldots,r}$ has determinant $1/\theta'^r$, so the joint density of $Y_1, \ldots, Y_r$ is

$$\begin{aligned}f_{Y_1,\ldots,Y_r}(y_1, \ldots, y_r) &= f_{Z_1,\ldots,Z_r}(y_1/\theta', (y_2 - y_1)/\theta', \ldots, (y_r - y_{r-1})/\theta')\frac{1}{\theta'} \\ &= \frac{1}{(2\theta')^r}\exp\left(-\frac{y_r}{2\theta'}\right), \qquad 0 \leq y_1 \leq y_2 \leq \cdots \leq y_r.\end{aligned}$$

Finally, since Y_r/θ' is the sum of r iid χ_2^2 distributions, it follows a χ_{2r}^2 distribution.

7. In statistics, a simple random sample is a subset of individuals chosen (one by one) from a population. Each individual is chosen randomly such that each individual has the same probability of being chosen at any stage during the sampling process, and each subset of k individuals has the same probability of being chosen for the sample as any other subset of k individuals.

   From a population of size N with finite variance, a simple random sample of size n is drawn without replacement, and a real-valued characteristic X measured to yield observations $X_j, j = 1, \ldots, n$.

   (a) Show that the sample mean $\bar{X}_n$ is an unbiased estimator of the population mean m.

Suppose the population with distinct feature is $y_1, y_2, \ldots, y_k$ with occurance time $c_1, c_2, \ldots, c_k$, so $c_1 + c_2 + \cdots + c_k = N$. The sample drawn without replacement is represented by $X_1, X_2, \ldots, X_n$, with replacement by $R_1, R_2, \ldots, R_n$. It is given that $\sum_{i=1}^k c_i y_i / \sum_{i=1}^k c_i = \sum_{i=1}^k c_i y_i / N = m$. Since each particular individual is equally likely to be selected in the sample, we can see that

$$\Pr(X_i = y_j) = \frac{c_j}{N}.$$

The expectation of X_i is therefore

$$\mathbb{E}(X_i) = \sum_{j=1}^k y_j \frac{c_j}{N} = m.$$

Hence the expectation of the sample mean is

$$\mathbb{E}(\bar{X}_n) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) = m.$$

(b) Show that the expected squared error of $\bar{X}_n$ as an estimator of m is smaller than that of the mean of a simple random sample of the same size n drawn with replacement.

Suppose the population variance is σ^2, i.e.

$$\frac{\sum_{i=1}^k c_i(y_i-m)^2}{\sum_{i=1}^k c_i} = \frac{\sum_{i=1}^k c_i(y_i-m)^2}{N} = \frac{\sum_{i=1}^k c_i y_i^2 - m^2 N}{N} = \sigma^2.$$

Then

$$\mathrm{Var}(X_i) = \mathbb{E}(X_i^2) - m^2 = \left(\sum_{j=1}^k y_j^2 \frac{c_j}{N}\right) - m^2 = \frac{\sum_{i=1}^k c_i y_i^2 - m^2 N}{N} = \sigma^2.$$

We now find the covariance between X_i and X_j for $i \neq j$ as follows: we can see that the joint density of X_i and X_j as

$$\Pr(X_i = y_s, X_j = y_t) = \frac{c_s c_t}{N(N-1)}, \qquad s \neq t,$$
$$\Pr(X_i = y_s, X_j = y_t) = \frac{c_s(c_t-1)}{N(N-1)}, \qquad s = t,$$

then for $i \neq j$,

$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= \mathbb{E}(X_i X_j) - m^2 \\
&= -m^2 + \sum_{s=1}^k \sum_{t=1}^k y_s y_t \Pr(X_i = y_s, X_j = y_t) \\
&= -m^2 + \sum_{s=1}^k \sum_{t=1}^k y_s y_t \Pr(X_j = y_t | X_i = y_t) \Pr(X_i = y_t) \\
&= -m^2 + \sum_{s=1}^k y_s \Pr(X_i = y_s) \sum_{t=1}^k y_t \Pr(X_j = y_t | X_i = y_s) \\
&= -m^2 + \sum_{s=1}^k y_s \Pr(X_i = y_s) \left[\sum_{t=1, t\neq s}^k y_t \frac{c_t}{N-1} + y_s \frac{c_s - 1}{N-1}\right] \\
&= -m^2 + \sum_{s=1}^k y_s \frac{c_s}{N} \left[\sum_{t=1}^k y_t \frac{c_t}{N-1} - y_s \frac{1}{N-1}\right] \\
&= -m^2 + \sum_{s=1}^k y_s \frac{c_s}{N} \left[\frac{mN}{N-1} - y_s \frac{1}{N-1}\right] \\
&= -m^2 + \sum_{s=1}^k y_s \frac{c_s}{N} \frac{mN}{N-1} - \sum_{s=1}^k y_s \frac{c_s}{N} y_s \frac{1}{N-1} \\
&= -m^2 + \frac{m^2 N}{N-1} - \frac{\sum_{s=1}^k c_s y_s^2}{N(N-1)} \\
&= -m^2 + \frac{m^2 N}{N-1} - \frac{\sigma^2 N + m^2 N}{N(N-1)} \\
&= -\frac{\sigma^2}{N-1},
\end{aligned}$$

so

$$\begin{aligned}
\mathrm{Var}(\bar{X}_n) &= \frac{1}{n^2}\left(\sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j)\right) \\
&= \frac{1}{n^2}\left(n\sigma^2 + -n(n-1)\frac{\sigma^2}{N-1}\right) \\
&= \left(1 - \frac{n-1}{N-1}\right)\frac{\sigma^2}{n}.
\end{aligned}$$

For simple random sample with replacement, we know that R_i's are iid with

$$\Pr(R_i = y_j) = \frac{c_j}{N}.$$

The expectation of R_i is therefore

$$\mathbb{E}(R_i) = \sum_{j=1}^{k} y_j \frac{c_j}{N} = m,$$

and variance is therefore

$$\mathrm{Var}(R_i) = \mathbb{E}(R_i^2) - m^2 = \left(\sum_{j=1}^{k} y_j^2 \frac{c_j}{N}\right) - m^2 = \frac{\sum_{i=1}^{k} c_i y_i^2 - m^2 N}{N} = \sigma^2.$$

So

$$\mathrm{Var}(\bar{R}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} R_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}.$$

Since the bias of these two estimators are the same, we know that the expected square error is equal to their variances, so

$$\frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) = \mathrm{MSE}(\bar{X}_n) < \mathrm{MSE}(\bar{R}_n) = \frac{\sigma^2}{n}.$$

(c) Show that as $n, N \to +\infty$ and $r = n/N \to 0$ and the population variance is always less than M for all N, the difference between the expected squared errors of the two estimators is $O(1/N)$.

$$\mathrm{MSE}(\bar{R}_n) - \mathrm{MSE}(\bar{X}_n) = \frac{n-1}{N-1}\frac{\sigma^2}{n} = O(1/N)$$

.