Solutions Chapter 2 (Estimation) – Mathematical Statistics

1. The solutions to a. and b. for the 4 estimators are as follows:

1. $E(T_1) = E(X_1) = \mu$ and $var(T_1) = var(X_1) = \sigma^2$, this is an anbiased estimator.
   So the Mean Squared Error is: $E(T_1 - \mu)^2 = var(X_1) = \sigma^2$
2. $E(T_2) = E\left(\frac{X_1+X_2}{2}\right) = \frac{1}{2}(EX_1 + EX_2) = \frac{1}{2}(\mu + \mu) = \mu$ (T_2 is an unbiased estimator!) and
   $var(T_2) = var\left(\frac{X_1+X_2}{2}\right) = \frac{1}{4}var(X_1 + X_2) = \frac{1}{4}\big(var(X_1) + var(X_2)\big) = \frac{1}{2}\sigma^2$, so
   $E(T_2 - \mu)^2 = var(T_2) = \frac{1}{2}\sigma^2$ ($< var(T_1)$, so T_2 is better than T_1)
3. $E(T_3) = EX_1 + EX_2 + \cdots + EX_{10} = 10\mu$ (not an unbiased estimator of μ) and
   (because of independence:) $var(T_3) = var(X_1) + \cdots + var(X_{10}) = 10\sigma^2$.
   Because of the bias we will use the formula $E(T - \theta)^2 = (ET - \theta)^2 + var(T)$:
   $E(T_3 - \mu)^2 = (ET_3 - \mu)^2 + var(T_3) = (10\mu - \mu)^2 + 10\sigma^2 = 10\sigma^2 + 81\mu^2$
4. $E(T_4) = E\left(\frac{X_1+X_2+\cdots+X_{10}}{10}\right) = \frac{EX_1+EX_2+\cdots+EX_{10}}{10} = \frac{10\mu}{10} = \mu$ (an unbiased estimator of μ) and
   $var(T_4) = var\left(\frac{X_1+X_2+\cdots+X_{10}}{10}\right) = \frac{1}{10^2}\big(var(X_1) + \cdots + var(X_{10})\big) = \frac{1}{100}\cdot 10\sigma^2 = \frac{1}{10}\sigma^2$.
   So $E(T_4 - \mu)^2 = var(T_4) = \frac{1}{10}\sigma^2$

In conclusion: the sample mean $T_4 = \overline{X}$ is the estimator on the basis of 10 variables, that has the smallest Mean Squared Error, and therefore it is the best of these four estimators.
(*Remark: T_1 and T_2 are sample means as well, but based on only 1 or 2 observations (larger variance), T_3 has a bias ànd a large variance.*)

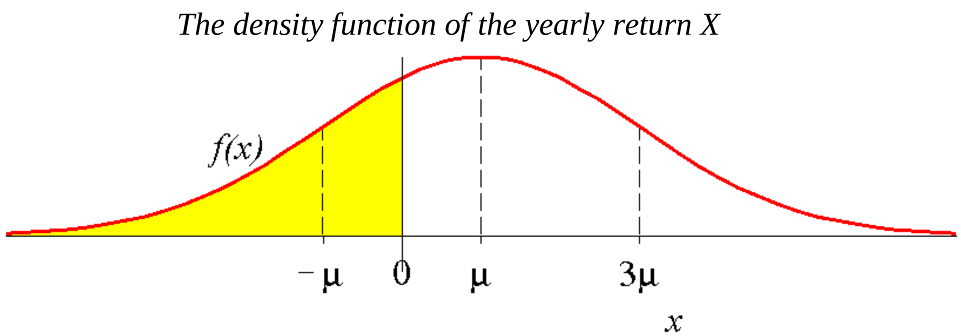
2.

a. If the estimator is unbiased if $E(T) = \mu$, which is true for T_1 and T_2:
$E(T_1) = E\left(\frac{1}{2}\left(\overline{X} + \overline{Y}\right)\right) = \frac{1}{2}\left(E\overline{X} + E\overline{Y}\right) = \frac{1}{2}(\mu + \mu) = \mu$ and $E(T_2) = \frac{mE(\overline{X})+nE(\overline{Y})}{m+n} = \frac{(m+n)\mu}{m+n} = \mu$

b. Since both estimators are unbiased, the best is the one with the smallest variance (= the Mean Squared Error for unbiased estimators):
$var(T_1) = \frac{1}{4}var\left(\overline{X} + \overline{Y}\right) = \frac{1}{4}\left(var(\overline{X}) + var(\overline{Y})\right) = \frac{1}{4}\left(\frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right) = \left(\frac{1}{4m} + \frac{1}{4n}\right)\sigma^2$ and
$var(T_2) = var\left(\frac{m}{m+n}\overline{X} + \frac{n}{m+n}\overline{Y}\right) = \frac{m^2}{(m+n)^2}var(\overline{X}) + \frac{n^2}{(m+n)^2}var(\overline{Y}) = \frac{m^2}{(m+n)^2}\cdot\frac{\sigma^2}{m} + \frac{n^2}{(m+n)^2}\cdot\frac{\sigma^2}{n}$
$= \left(\frac{m}{(m+n)^2} + \frac{n}{(m+n)^2}\right)\sigma^2 = \frac{m+n}{(m+n)^2}\sigma^2 = \frac{1}{m+n}\sigma^2$
$var(T_1) > var(T_2)$ if $\frac{1}{4m} + \frac{1}{4n} > \frac{1}{m+n}$ or: $\frac{(m+n)n+(m+n)m-4mn}{4(m+n)mn} = \frac{n^2+m^2-2mn}{4(m+n)mn} < 0$,
so if $n^2 + m^2 - 2mn = (m - n)^2 > 0$. This inequality hols for all $m \neq n$.
So T_2 is the best estimator of μ, except if $m = n$.
In the latter case they are equally good (then $T_2 = T_1$).

3.

a.



$P(X<0) = P\left[\frac{X-\mu}{2\mu} < \frac{0-\mu}{2\mu}\right] = P(Z < -½) = 1 - \Phi(0.50) = 30.85\%$ $(Z \sim N(0,1))$

b. $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and since $\sigma^2 = 4\mu^2$, we have $\overline{X} \sim N\left(\mu, \frac{4\mu^2}{n}\right)$

c. Y is an unbiased estimator of μ if $E(Y) = \mu$.
For the sample mean $E(\overline{X})$ μ is always true.
The unbiasedness of Y implies $E(Y) = $ μ:
$E(Y) = E(a\overline{X}) = aE(\overline{X}) = a\mu = \mu$, so $a = 1$.

d. For Y we have: $E(Y) = a\mu$, (*is not equal to μ if a* ≠1)
and $var(Y) = var(a\overline{X}) = a^2\, var(\overline{X}) = a^2 \times \frac{4\mu^2}{n} = \frac{4a^2\mu^2}{10} = \frac{2}{5}a^2\mu^2$ (*if a decreases, var(Y) decreases*)
*(Note that E(Y - μ)² = (**E(Y) - μ**)² + **var(Y)** really splits the Mean Squared Error into the **bias** E(Y) - μ and the **variance of Y**: if we choose a smaller than 1, the second term **var(Y)** will decrease, but on the other hand the first term (**E(Y) - μ**)² will increase (larger than 0). We are searching the value of a such that the sum of these two effects gives a minimum.)*
The Mean Squared Error is, expressed in unknown (fixed) μ and a:

$$E(Y-\mu)^2 = (E(Y)-\mu)^2 + var(Y) = (a\mu-\mu)^2 + \frac{2a^2\mu^2}{5} = \mu^2\left[(a-1)^2 + \frac{2a^2}{5}\right]$$

This Mean Squared Error takes on its smallest value if the value $f(a) = (a-1)^2 + \frac{2a^2}{5}$ is minimal (assuming that μ ≠ 0).
The derivative is 0 in the extreme values of f: $f\,`(a) = 2a - 2 + \frac{4a}{5} = 0$, or: $2.8a = 2$: $\boldsymbol{a} = \frac{\mathbf{5}}{\mathbf{7}}$
Considering the signs of the derivative (or computing the second derivative f ``(5/7) = 2.8 > 0), we can draw the conclusion that f attains its minimum at this value.
$Y = \frac{5}{7}\overline{X} = \frac{1}{14}\sum_{i=1}^{10} X_i$ is the best estimator of μ: $E(Y) = \frac{5}{7}$μ and $var(Y) = \frac{2a^2\mu^2}{5} = \frac{10\mu^2}{49}$.

4. Remember that the maximum likelihood method chooses the value of the population parameter, for which the observed outcome of the experiment is most likely.
The outcome in this case is "one red and one white" when drawing 2 marbles arbitrarily from either 3 red + 3 white marbles or from 2 red + 2 white marbles.

$$P(1 \text{ red and } 1 \text{ white}|\text{total} = 6) = \frac{\binom{3}{1}\binom{3}{1}}{\binom{6}{2}} = \frac{9}{15} = 60\%$$

and $P(1 \text{ red and } 1 \text{ white}|\text{total} = 4) = \frac{\binom{2}{1}\binom{2}{1}}{\binom{4}{2}} = \frac{4}{6} \approx 66.7\%$

Since the last probability is the largest, 4 marbles (2 red and 2 white) is the maximum likelihood estimate of the total number of the marbles.

5.

a. Likelihood function (as a function of μ, given the observations $x_1, \ldots, x_n$):

$$L(\mu) = P(X_1 = x_1) \times \ldots . \times P(X_n = x_n) = \prod_i \frac{\mu^{x_i}}{x_i!}e^{-\mu} = \frac{\mu^{x_1+\cdots+x_n}}{x_1! \cdot \ldots \cdot x_n!}e^{-n\mu},\ \mu > 0$$

log likelihood: $lnL(\mu) = (x_1 + \cdots + x_n)\ln\mu - n\mu - \ln(x_1! \cdot \ldots \cdot x_n!)$, with $\mu > 0$

$\frac{d}{d\mu}lnL(\mu) = \frac{x_1+\cdots+x_n}{\mu} - n = 0$, if $\mu = \frac{x_1+\cdots+x_n}{n} = \overline{x}$.

$lnL(\overline{x}) =$ is a maximum, since $\frac{d^2}{d\mu^2}lnL(\mu) = -\frac{x_1+\cdots+x_n}{\mu^2} < 0$ $(x_i \geq 0)$

(or since $\frac{d}{d\mu}lnL(\mu) = \frac{x_1+\cdots+x_n}{\mu} - n < 0$ *if* $\mu > \overline{x}$ and > 0 if $\mu > \overline{x}$)

Conclusion: $\hat{\mu} = \overline{X}$ is the maximum likelihood estimator of μ.

The sample mean $\overline{X}$ is always an unbiased and consistent estimator of μ if σ^2 exists: $E(\overline{X}) = \mu$ and $var(\overline{X}) = \frac{\sigma^2}{n} \to 0$ if $n \to \infty$, so that $MSE(\overline{X}) = (E\overline{X} - \mu)^2 + var(\overline{X}) \to 0$ as $n \to \infty$.

By definition, $V = \overline{X}$ is a sufficient estimator **if** the conditional distribution of $X_1, X_2, \ldots X_n$, given $V = v$, does not depend on μ (μ > 0). *(We will first use the* *<u>definition</u>* *of sufficiency straight away)*

Well, we have $P(X_1 = x_1 \text{ and } \ldots \text{ and } X_n = x_n) = L(\mu) = \frac{\mu^{x_1+\cdots+x_n}}{x_1!\cdot\ldots\cdot x_n!}e^{-n\mu}$,

$V = \overline{X} = v$ is equivalent with $\overline{X} = v$ or $X_1 + \cdots + X_n = n \cdot v$

For independent and Poisson distributed variables the sum is Poisson distributed as well, with parameter $n\mu$, so:

$$\begin{aligned} P(X_1 = x_1, \ldots, X_n = x_n \mid V = v) &= \frac{P(X_1 = x_1, \ldots, X_2 = x_n, V = v)}{P(V = v)} \\ &= \frac{P(X_1 = x_1 \text{ and } \ldots \text{ and } X_n = x_n \text{ and } x_1 + \cdots + x_n = nv)}{P(X_1 + \cdots + X_n = nv)} \\ &= \frac{\frac{\mu^{nv}}{x_1! \cdot \ldots \cdot x_n!}e^{-n\mu}}{\frac{(n\mu)^{nv}}{(nv)!}e^{-n\mu}} = \frac{(nv)!}{x_1! \cdot \ldots \cdot x_n! \; n^{nv}} \quad (nv = 0,1,2,3,\ldots) \end{aligned}$$

This conditional distribution does not depend on μ, so $\overline{X}$ is a sufficient estimator for μ.

Using Fisher-Neyman factorization (see the theorem on the slides of week 2) we find:

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = L(\mu) = \frac{\mu^{x_1+\cdots+x_n}}{x_1! \cdot \ldots \cdot x_n!}e^{-n\mu} = \frac{1}{x_1! \cdot \ldots \cdot x_n!} \cdot [\, \mu^{x_1+\cdots+x_n}e^{-n\mu}],$$

where $h(x_1, \ldots, x_n) = \frac{1}{x_1!\cdot\ldots\cdot x_n!}$ does not depend on μ and $g(x_1 + \cdots + x_n) = \mu^{x_1+\cdots+x_n}e^{-n\mu}$ or $g(\overline{x}) = \mu^{n\overline{x}}e^{-n\mu}$. Hence $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ is a sufficient estimator of μ.

b. Likelihood function (as a function of the unknown p):

$$\begin{aligned} L(p) = P(X_1 = x_1) \times \ldots \times P(X_n = x_n) &= \prod_i (1-p)^{x_i - 1}p \\ &= (1-p)^{x_1+\cdots+x_n-n}p^n \end{aligned}$$

log likelihood: $lnL(p) = (x_1 + \cdots + x_n - n)\ln(1-p) + n \cdot ln(p)$, where $0 < p < 1$

$$\frac{d}{dp}lnL(p) = -\frac{x_1+\cdots+x_n-n}{1-p} + \frac{n}{p} = \frac{-(x_1+\cdots+x_n-n)p - n(1-p)}{(1-p)p} = \frac{n-(x_1+\cdots+x_n)p}{(1-p)p} = 0,$$

if $n = (x_1 + \cdots + x_n)p$ or $\hat{p} = \frac{n}{x_1+\cdots+x_n} = {}^1\!/_{\overline{x}}$

$lnL\left({}^1\!/_{\overline{x}}\right)$ is a maximum, since $\frac{d}{dp}lnL(p) = \frac{n-(x_1+\cdots+x_n)p}{(1-p)p} < 0$ if $p > {}^1\!/_{\overline{x}}$

and $\frac{d}{dp}lnL(p) > 0$ if $p < {}^1\!/_{\overline{x}}$)

Conclusion: ${}^1\!/_{\overline{X}}$ is the maximum likelihood estimator of p.

For $n = 2$, by definition $V = {}^1\!/_{\overline{X}}$ is a sufficient estimator **if** the conditional distribution of X_1 and X_2, given $V = v$, does not depend on p $(0 < p < 1)$.

Well, we have $P(X_1 = x_1 \text{ and } X_2 = x_2) = L(p) = (1-p)^{x_1+x_2-2}p^2$.

$V = {}^1\!/_{\overline{X}} = v$ is equivalent with $\overline{X} = \frac{1}{v}$ or $X_1 + X_2 = \frac{2}{v}$.

$$\begin{aligned} P(X_1 + X_2 = z) &\overset{\text{convolution sum}}{=} \sum_{x_1=1}^{z-1} P(X_1 = x_1)P(X_2 = z - x_1) \\ &= \sum_{x_1=1}^{z-1} (1-p)^{x_1+(z-x_1)-2}p^2 \\ &= (z-1)(1-p)^{z-2}p^2, \; z = 2,3,4,\ldots \text{ (negative binomial distribution)} \end{aligned}$$

So

$$P(X_1 = x_1 \text{ and } X_2 = x_2 \mid V = v) = \frac{P(X_1 = x_1 \text{ and } X_2 = x_2 \text{ and } V = v)}{P(V = v)}$$

$$= \frac{P\left(X_1 = x_1 \text{ and } X_2 = x_2 \text{ and } x_1 + x_2 = \frac{2}{v}\right)}{P\left(X_1 + X_2 = \frac{2}{v}\right)} = \frac{(1-p)^{\frac{2}{v}-2}p^2}{\left(\frac{2}{v}-1\right)(1-p)^{\frac{2}{v}-2}p^2} = \frac{1}{\frac{2}{v}-1} \quad \left(\frac{2}{v} = 2, 3, 4, \ldots.\right)$$

This conditional distribution does not depend on p, so $\overline{X}$ is a sufficient estimator (*general case is more complicated since we need the convolution of* $X_1, \ldots, X_n$, *given by the negative binomial distribution*).
(applying Fisher-Neyman factorization:
the joint probability function $P(X_1 = x_1$ and $X_2 = x_2) = (1-p)^{x_1+x_2-2}p^2$ is a function of $x_1 + x_2$.
Since $1/\overline{X} = \frac{2}{X_1+X_2}$, we have $g\left(1/\overline{x}\right) = (1-p)^{2\overline{x}}p^2 = (1-p)^{x_1+x_2-2}p^2$ (*a function of* $1/\overline{x}$: $\overline{x} = \frac{1}{1/\overline{x}}$)
Hence $1/\overline{X}$ is a sufficient estimator of p.

c. (*Note that* σ^2 *(not* σ*) is the parameter in this case, so L is a function of* $\sigma^2 = \theta$)

Likelihood function $L(\sigma^2) = f_{X_1}(x_1) \times \ldots.\times f_{X_n}(x_n) = \prod_i (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{(x_i-10)^2}{2\sigma^2}}$

$$= (2\pi\sigma^2)^{-\frac{1}{2}n} e^{-\frac{\sum_{i=1}^n (x_i-10)^2}{2\sigma^2}}, \ \sigma^2 > 0$$

log likelihood: $lnL(\sigma^2) = -\frac{1}{2}n\ln(2\pi\sigma^2) - \frac{\sum_{i=1}^n (x_i-10)^2}{2\sigma^2}$, with $\sigma^2 > 0$ (*you might substitute* $z = \sigma^2$)

$$\frac{d}{d(\sigma^2)} lnL(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i-10)^2}{2(\sigma^2)^2} = \frac{1}{2\sigma^4}\left[\sum_{i=1}^n (x_i-10)^2 - n\sigma^2\right] = 0,$$

if $\sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i-10)^2$.

$lnL(\overline{x}) =$ is a maximum, since $\frac{d}{d(\sigma^2)} lnL(\sigma^2) < 0$ if $\sigma^2 > \frac{1}{n}\sum_{i=1}^n (x_i-10)^2$ and
$\frac{d}{d(\sigma^2)} lnL(\sigma^2) > 0$ if $\sigma^2 < \frac{1}{n}\sum_{i=1}^n (x_i-10)^2$.

Conclusion: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - 10)^2$ is the mle of σ^2.

It is unbiased and consistent since $E\left[\frac{1}{n}\sum_{i=1}^n (X_i-10)^2\right] = \frac{1}{n}\sum_{i=1}^n E(X_i-10)^2 = \frac{1}{n}\cdot n\cdot\sigma^2 = \sigma^2$ and

$$\lim_{n\to\infty} var\left[\frac{1}{n}\sum_{i=1}^n (X_i-10)^2\right] = \lim_{n\to\infty} \frac{1}{n^2}\left[\sum_{i=1}^n var(X_i-10)^2\right] = \lim_{n\to\infty} \frac{var(X-10)^2}{n} = 0.$$

(if the variance exists, which is the case for a normal distribution, You might also use the fact that for the sample variance for known μ, $\frac{nS_\mu^2}{\sigma^2}$ has a χ_n^2-distribution: $var\left(\frac{nS_\mu^2}{\sigma^2}\right) = 2n \implies var(S_\mu^2) = \cdots$.).

6.

a. $\frac{(n-1)S^2}{\sigma^2}$ has a χ^2-distribution with $n-1$ degrees of freedom and the χ_{n-1}^2-distribution has an expectation $n-1$ and a variance $2(n-1)$, see also exercise 12.
So $var\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1)$ or $\frac{(n-1)^2}{\sigma^4} var(S^2) = 2(n-1)$ or: $var(S^2) = \frac{2\sigma^4}{n-1}$.

b. No, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2 = \frac{n-1}{n}S^2$ (not an unbiased estimator)
So $E(\hat{\sigma}^2) = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}E(S^2) = \frac{n-1}{n}\sigma^2$
And: $var(\hat{\sigma}^2) = var\left(\frac{n-1}{n}S^2\right) = \frac{(n-1)^2}{n^2}var(S^2) = \frac{2(n-1)}{n^2}\sigma^4$.
$MSE(S^2) = var(S^2) = \frac{2\sigma^4}{n-1}$
$MSE(\hat{\sigma}^2) = (E(\hat{\sigma}^2) - \sigma^2)^2 + var(\hat{\sigma}^2) = \frac{\sigma^4}{n^2} + \frac{2(n-1)}{n^2}\sigma^4 = \frac{2n-1}{n^2}\cdot\sigma^4$
$\frac{2n-1}{n^2} < \frac{2}{n-1}$, since $(2n-1)(n-1) < 2n^2 \Leftrightarrow 2n^2 - 2n - n + 1 < 2n^2 \Leftrightarrow 1 - 3n < 0$, which is true

for all $n = 2, 3, ...$ (for $n = 1$ S^2 is not defined and $\hat{\sigma}^2 = 0$)

So $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is a better estimator than $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

c. The best estimator of the "family" $T = a \cdot S^2$ (a in $\mathbb{R}$).

$E(aS^2) = a\sigma^2$ and $var(aS^2) = a^2 var(S^2) = \frac{2a^2\sigma^4}{n-1}$

Then $MSE(T) = (T - \sigma^2)^2 + var(T) = (a-1)^2\sigma^4 + \frac{2a^2\sigma^4}{n-1} = \left[(a-1)^2 + \frac{2a^2}{n-1}\right]\sigma^4$ is minimal if

$f(a) = (a-1)^2 + \frac{2a^2}{n-1}$ is minimal.

$f`(a) = 2(a+1) + \frac{4a}{n-1} = 0$, if $\left(1 + \frac{2}{n-1}\right)a = 1$, so if $a = \frac{n-1}{n+1}$.

Since $f``(a) = 2 + \frac{4}{n-1} > 0$, f is minimal for $a = \frac{n-1}{n+1}$.

Conclusion: $T = \frac{n-1}{n+1} \cdot S^2 = \frac{1}{n+1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is the best estimator of all $T = a \cdot S^2$.

7. **a.** X = "the weight of an arbitrary egg" ~ $N(\mu, \sigma^2)$, so $P(X > 68.5) = P\left(Z > \frac{68.5-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{68.5-\mu}{\sigma}\right)$

b. The mle of the probability $P(X > 68.5)$ is $1 - \Phi\left(\frac{68.5-\hat{\mu}}{\hat{\sigma}}\right)$,

where $\hat{\mu} = \overline{x} = 56.3$ gram, but $\hat{\sigma} = \sqrt{\frac{1}{n}\sum(x_i - \overline{x})^2}$ (see the theory) is not directly given:

Since $s = \sqrt{\frac{1}{n-1}\sum(x_i - \overline{x})^2} = 7.6$ and $n = 25$, we have $\hat{\sigma}^2 = \frac{24s^2}{25} = \frac{24 \cdot 7.6^2}{25}$, so $\hat{\sigma} = \sqrt{\frac{24 \cdot 7.6^2}{25}} \approx 7.45$.

The mle of $P(X > 68.5)$ is $1 - \Phi\left(\frac{68.5-56.3}{7.45}\right) \approx 1 - \Phi(1.64) = 0.0505$.

8. **a**. $T = aT_1 + bT_2 + cT_3$ is unbiased if $E(T) = E(aT_1 + bT_2 + cT_3) = \mu$

$E(aT_1 + bT_2 + cT_3) = aE(T_1) + bE(T_2) + cE(T_3) = (a + b + c)\mu = \mu$ if $a + b + c = 1$.

$var(aT_1 + bT_2 + cT_3) = a^2 var(T_1) + b^2 var(T_2) + c^2 var(T_3) = (a^2 + 2b^2 + 3c^2)\sigma^2$,

where $c = 1 - a - b$, so minimize the quadratic function $f(a, b) = a^2 + 2b^2 + 3(1 - a - b)^2$:

$\frac{\partial}{\partial a} f(a, b) = 2a - 6(1 - a - b) = 8a + 6b - 6 = 0$ and $\frac{\partial}{\partial b} f(a, b) = 10b + 6a - 6 = 0$.

Solving the equations we find $a = \frac{6}{11}$ and $b = \frac{3}{11}$ (and $c = \frac{2}{11}$)

$f\left(\frac{6}{11}, \frac{3}{11}\right)$ is a minimum,

because it is the only extreme of the positive f, that does not have an upper bound.

b. $\overline{X}$ has a variance $\frac{\sigma^2}{n}$: $\overline{X}$ is not the unbiased linear combination of the X_i's with the smallest variance,

there must be another estimator T with coefficients $a_i = \frac{1}{n} + \delta_i$, such that $\sum \delta_i = 0$ ($\sum a_i = 1$).

But then $var(T) = var\left[\sum_{i=1}^{n}\left(\frac{1}{n} + \delta_i\right)X_i\right] = \sigma^2 \sum_{i=1}^{n}\left(\frac{1}{n} + \delta_i\right)^2$

$\sum_{i=1}^{n}\left(\frac{1}{n} + \delta_i\right)^2 = \sum_{i=1}^{n}\left(\frac{1}{n^2} + \frac{2\delta_i}{n} + \delta_i^2\right) = \frac{1}{n} + \frac{2}{n} \cdot \sum_{i=1}^{n}\delta_i + \sum_{i=1}^{n}\delta_i^2 = \frac{1}{n} + \sum_{i=1}^{n}\delta_i^2 \geq \frac{1}{n}$,

$var(T) \geq \frac{\sigma^2}{n} = var(\overline{X})$, contradicting the assumption that T has a smaller variance: $\overline{X}$ is the best.

9.

a. $f(x) = \frac{1}{\theta}$ for $0 \leq x \leq \theta$, $F(x) = \frac{x}{\theta}$ for $0 \leq x \leq \theta$, $E(X) = \frac{\theta}{2}$ (symmetry) and $var(X) = \frac{\theta^2}{12}$.

b. Likelihood function $(\theta) = f_{X_1}(x_1) \times .... \times f_{X_n}(x_n) = \prod_i \left(\frac{1}{\theta}\right) = \frac{1}{\theta^n}$,

with $0 < x_i \leq \theta$ for all $i = 1, ..., n$, so $\theta \geq \max(x_1, ..., x_n)$

$L(\theta)$ is a decreasing function ($L`(\theta) = -\frac{n}{\theta^{n+1}} < 0$), so $L(\theta)$ attains its maximum at the bound of its domain: $\theta = \max(x_1, \dots, x_n)$. Therefore the *mle* is $\hat{\theta} = \max(X_1, \dots, X_n) = T_1$.

c. First we need the density function of $\hat{\theta} = \max(X_1, \dots, X_n)$:

$$F_{T_1}(x) = P(\max(X_1, \dots, X_n) \le x) = P(X_1 \le x) \cdot \dots \cdot P(X_n \le x) = \left(\frac{x}{\theta}\right)^n, 0 \le x \le \theta$$

$$f_{T_1}(x) = \frac{d}{dx} F_{T_1}(x) = \frac{n}{\theta} \cdot \left(\frac{x}{\theta}\right)^{n-1}, 0 \le x \le \theta$$

$$E(T_1) = \int_{-\infty}^{\infty} x\, f_{T_1}(x) dx = \int_0^{\theta} x \cdot \frac{n}{\theta} \cdot \left(\frac{x}{\theta}\right)^{n-1} dx = \left[\frac{n}{n+1} \cdot \frac{x^{n+1}}{\theta^n}\right]_0^{\theta} = \frac{n}{n+1}\theta \quad \neq \theta$$

So T_1 is not unbiased, but $aT_1 = \frac{n+1}{n} T_1$ is an unbiased estimator: $E\left(\frac{n+1}{n} T_1\right) = \frac{n+1}{n} E(T_1) = \theta$.

You could verify that T_1 is a better estimator than $\frac{n+1}{n} T_1$, since its MSE is smaller (use this result in d.)

d. $MSE(T_1) = (ET_1 - \theta)^2 + var(T_1)$

Using the result in c.: $(ET_1 - \theta)^2 = \frac{\theta^2}{(n+1)^2}$.

$$E(T_1^2) = \int_{-\infty}^{\infty} x^2 f_{T_1}(x) dx = \int_0^{\theta} x^2 \cdot \frac{nx^{n-1}}{\theta^n} dx = \left[\frac{n}{n+2} \cdot \frac{x^{n+2}}{\theta^n}\right]_0^{\theta} = \frac{n}{n+2}\theta^2$$

$$var(T_1) = E(T_1^2) - (ET_1)^2 = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{(n+2)(n+1)^2}\theta^2$$

$$MSE(T_1) = (ET_1 - \theta)^2 + var(T_1) = \frac{\theta^2}{(n+1)^2} + \frac{n}{(n+2)(n+1)^2}\theta^2 = \frac{2n+2}{(n+2)(n+1)^2}\theta^2$$

So $\lim_{n\to\infty} MSE(T_1) = 0$: the *mle* T_1 is a consistent estimator of θ.

e. $E(T_2) = E(2\overline{X}) = 2E(\overline{X}) = 2\mu = 2 \cdot \frac{1}{2}\theta = 0$: T_2 is unbiased.

f. $MSE(T_1) = \frac{2n+1}{(n+2)(n+1)^2}\theta^2$

Since T_2 is unbiased, $MSE(T_2) = var(T_2) = var(2\overline{X}) = 4var(\overline{X}) = 4 \cdot \frac{\frac{\theta^2}{12}}{n} = \frac{\theta^2}{3n}$.

$MSE(T_1) < MSE(T_2)$ if $\frac{2n+1}{(n+2)(n+1)^2}\theta^2 < \frac{\theta^2}{3n}$ or $(2n+1)3n < (n+2)(n+1)^2$ $(\theta > 0)$

or $[n^3 + 4n^2 + 5n + 2] - [6n^2 + 3n] > 0$ or $n^3 - 2n^2 + 2n + 2 > 0$.

This cubic inequality is not easily solved, but, since n is an integer, we can reason:

$n^3 - 2n^2 + 2n + 2 = n(n-2) + 2(n+1) > 0$ for all $n = 2, 3, 4, \dots$ and for $n = 1$: $-1 + 4 > 0$

Conclusion: T_1 is better than T_2 for all $n = 1, 2, \dots$ (using the MSE-criterion)

Note that T_1, the biased estimator, is better than the unbiased T_2.

The main difference is the variance: T_1 has a variance (and MSE), that approaches 0 with a factor $\frac{1}{n^2}$, where the variance of T_2 has a factor $\frac{1}{n}$.

10. Note that in this exercise we do not have independent variables, but a joint distribution with unknown parameters p_1 and p_2.

$L(p_1, p_2) = P(X_1 = x_1 \text{ and } X_2 = x_2 \text{ and } X_3 = x_3) = \frac{n!}{x_1! x_2! (n-x_1-x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n-x_1-x_2}$,

where $0 < p_1 < 1$, $0 < p_2 < 1$ and $0 < p_1 + p_2 < 1$.

$\ln L(p_1, p_2) = c + x_1 \ln(p_1) + x_2 \ln(p_2) + (n - x_1 - x_2) \ln(1 - p_1 - p_2)$

$\frac{\partial}{\partial p_1} \ln L(p_1, p_2) = \frac{x_1}{p_1} - \frac{n-x_1-x_2}{1-p_1-p_2} = 0$ and $\frac{\partial}{\partial p_2} \ln L(p_1, p_2) = \frac{x_2}{p_2} - \frac{n-x_1-x_2}{1-p_1-p_2} = 0$

Solving these equations we find $p_1 = \frac{x_1}{n}$ and $p_2 = \frac{x_2}{n}$.

This solution for the pair (p_1, p_2) must be a maximum, since the continuous function $L(p_1, p_2) \ge 0$ and on the bounds, such as $p_1 = 0$, $L(p_1, p_2) = 0$: if there is one extreme value, it must be a maximum.

So $\hat{\theta} = (\hat{p}_1, \hat{p}_2) = \left(\frac{X_1}{n}, \frac{X_2}{n}\right)$ is the mle of the pair (p_1, p_2).

$\hat{p}_1$ and $\hat{p}_2$ are consistent estimators: $\hat{p}_1 = \frac{X_1}{n}$, where X_1 is $N(n, p_1)$, so $E(\hat{p}_1) = \frac{E(X_1)}{n} = \frac{np_1}{n} = p_1$ and $var(\hat{p}_1) = \frac{var(X_1)}{n^2} = \frac{np_1(1-p_1)}{n^2} = \frac{p_1(1-p_1)}{n} \to 0$ if $n \to \infty$, so $\hat{p}_1$ is an unbiased and consistent estimator of p_1, and $\hat{p}_2$ likewise.

11. We consider two independent samples with the same sample size n:
$X_1, \ldots, X_n$ is a random sample of X, that has a $N(\mu_1, \sigma^2)$-distribution
$Y_1, \ldots, Y_n$ is a random sample of Y, that has a $N(\mu_2, \sigma^2)$-distribution (equal variances).
The parameters μ_1, μ_2 and σ^2

a. $L(\mu_1, \mu_2, \sigma^2) = f_{X_1}(x_1) \times \ldots . \times f_{X_n}(x_n) \times f_{X_1}(x_1) \times \ldots . \times f_{X_n}(x_n)$

$$= \prod_i (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{(x_i-\mu_1)^2}{2\sigma^2}} \times \prod_j (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{(y_j-\mu_2)^2}{2\sigma^2}}$$

$$= (2\pi\sigma^2)^{-\frac{1}{2}n-\frac{1}{2}n} e^{-\frac{\sum_{i=1}^n (x_i-\mu_1)^2 + \sum_{j=1}^n (y_j-\mu_2)^2}{2\sigma^2}} \quad \sigma^2 > 0$$

log likelihood: $lnL(\mu_1, \mu_2, \sigma^2) = -n\ln(2\pi\sigma^2) - \frac{\sum_{i=1}^n (x_i-\mu_1)^2 + \sum_{j=1}^n (y_j-\mu_2)^2}{2\sigma^2}$, with $\sigma^2 > 0$

* $\frac{\partial}{\partial \mu_1} lnL(\mu_1, \mu_2, \sigma^2) = \frac{\sum_{i=1}^n 2(x_i-\mu_1)}{2\sigma^2} = 0$, if $\mu_1 = \frac{1}{n}\sum_{i=1}^n x_i = \overline{x}$
* $\frac{\partial}{\partial \mu_2} lnL(\mu_1, \mu_2, \sigma^2) = \frac{\sum_{i=1}^n 2(y_i-\mu_2)}{2\sigma^2} = 0$, if $\mu_2 = \frac{1}{n}\sum_{i=1}^n y_i = \overline{y}$

For fixed σ^2 , $lnL(\mu_1, \mu_2, \sigma^2)$ attains its maximum value at $(\mu_1, \mu_2) = (\overline{x}, \overline{y})$
and for fixed (μ_1, μ_2):

* $\frac{\partial}{\partial(\sigma^2)} lnL(\overline{x}, \overline{y}, \sigma^2) = -\frac{n}{\sigma^2} + \frac{\sum_{i=1}^n (x_i-\overline{x})^2 + \sum_{j=1}^n (y_j-\overline{y})^2}{2(\sigma^2)^2}$

$$= \frac{1}{2\sigma^4}\left[\sum_{i=1}^n (x_i-\overline{x})^2 + \sum_{j=1}^n (y_j-\overline{y})^2 - 2n\sigma^2\right] = 0,$$

if $\sigma^2 = \frac{1}{2n}\left[\sum_{i=1}^n (x_i-\overline{x})^2 + \sum_{j=1}^n (y_j-\overline{y})^2\right]$, which is a maximum considering the signs of the derivative.

The maximum likelihood estimator of σ^2 is $\hat{\sigma}^2 = \frac{1}{2n}\left[\sum_{i=1}^n (X_i-\overline{X})^2 + \sum_{j=1}^n (Y_j-\overline{Y})^2\right]$.

b. Remember that $\frac{(n-1)S_X^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i-\overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}$, having an expectation $n-1$ and a variance $2(n-1)$:

we know that $E\left(\frac{(n-1)S_X^2}{\sigma^2}\right) = E\left(\frac{\sum_{i=1}^n (X_i-\overline{X})^2}{\sigma^2}\right) = n-1$, so $E\left(\sum_{i=1}^n (X_i-\overline{X})^2\right) = (n-1)\sigma^2$

and $var\left(\frac{(n-1)S_X^2}{\sigma^2}\right) = var\left(\frac{\sum_{i=1}^n (X_i-\overline{X})^2}{\sigma^2}\right) = 2(n-1)$, so $var\left(\sum_{i=1}^n (X_i-\overline{X})^2\right) = 2(n-1)\sigma^4$

$$E(\hat{\sigma}^2) = \frac{1}{2n}E\left[\sum_{i=1}^n (X_i-\overline{X})^2\right] + \frac{1}{2n}E\left[\sum_{j=1}^n (Y_j-\overline{Y})^2\right] = \frac{1}{2n}\cdot(n-1)\sigma^2 + \frac{1}{2n}\cdot(n-1)\sigma^2$$

$= \frac{n-1}{n}\sigma^2$ ($\hat{\sigma}^2$ is not unbiased, but asymptotically unbiased: $\lim_{n\to\infty}\frac{n-1}{n} = 1$).

Then $var(\hat{\sigma}^2) = \left(\frac{1}{2n}\right)^2\left[var\left(\sum_{i=1}^n (X_i-\overline{X})^2\right) + var\left(\sum_{j=1}^n (Y_j-\overline{Y})^2\right)\right]$

$= \frac{2\cdot 2(n-1)\sigma^4}{4n^2} = \frac{n-1}{n^2}\sigma^4 \to \infty$ if $n \to \infty$: $\widehat{\sigma^2}$ is a consistent estimator

c. If $n_1 = n_2 = n$, the pooled sample variance is $S^2 = \frac{1}{2(n-1)}\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{j=1}^{n}(Y_j - \overline{Y})^2\right]$

The factor $\frac{1}{2(n-1)}$ makes S^2 an unbiased estimator of σ^2.

12. $Z_1, Z_2, \ldots, Z_n$ are independent and all standard normal.

a. $E(Z_1^2) = var(Z_1) + (EZ_1)^2 = 1 + 0 = 1,$

$$E(Z_1^4) = \int_{-\infty}^{\infty} z^4 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz = \left[z^3 \cdot -\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\right]_{-\infty}^{-\infty} + \int_{-\infty}^{\infty} 3z^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz = 0 + 3E(Z_1^2) = 3$$

$var(Z_1^2) = E(Z_1^4) - E(Z_1^2)^2 \;\; = 3 - 1^2 = 2,$

$E(Z_1^2 + \cdots + Z_n^2) = n \cdot E(Z_1^2) = n$ and

$var(Z_1^2 + \cdots + Z_n^2) = n \cdot var(Z_1^2) = 2n$

b. (Below $Z = Z_1$ and $Y = Z_1^2$)

1. $F_Y(y) = P(Y \le y) = P(Z^2 \le y) = \begin{cases} 0 & \text{if } y \le 0 \\ P(-\sqrt{y} \le Z \le \sqrt{y}) & \text{if } y > 0 \end{cases}$

So, if $y > 0$: $F_Y(y) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y})$

2. If $y \le 0$, then $f_Y(y) = \frac{d}{dy}F_Y(y) = 0$

If $y > 0$, then $f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}[\Phi(\sqrt{y}) - \Phi(-\sqrt{y})]$

$$= \frac{1}{2\sqrt{y}}[\varphi(\sqrt{y}) + \varphi(-\sqrt{y})]$$

3. $f_Y(y) = 0,$ for $y \le 0$ and

$$f_Y(y) = \frac{1}{2\sqrt{y}}\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y})^2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y})^2}\right] = \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}y}, \qquad \text{for } y > 0.$$

c. $f_{X_1+X_2}(z) = \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(z-x)\,dx \; = \int_0^z \frac{1}{\sqrt{2\pi x}} e^{-\frac{1}{2}x} \cdot \frac{1}{\sqrt{2\pi x}} e^{-\frac{1}{2}(z-x)}\mathrm{dx} = \frac{e^{-\frac{1}{2}z}}{2\pi} \int_0^z \frac{1}{\sqrt{x(z-x)}}\,dx$

The last integral equals π (given): $f_{X_1+X_2}(z) = \frac{1}{2} e^{-\frac{1}{2}z}$, for $z > 0$.

A Chi-square distribution with 2 degrees of freedom is apparently the same as an Exp$\left(\frac{1}{2}\right)$*-distribution*

d. $f(z) = \frac{z^{\alpha-1} e^{-\frac{z}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}} = \frac{1 \cdot e^{-z/2}}{1 \cdot 2^1} = \frac{1}{2} e^{-\frac{1}{2}z}.$

13. The exercise is stated wrongly. The questions should be: use **Markov`s** inequality ……..

Solution:

Markov's inequality: for any random variable X and number $\varepsilon > 0$ we have: $P(|X| > \varepsilon) \le \frac{E(X^2)}{\varepsilon^2}$.

substituting $X = T - \theta$, we obtain the desirable inequality, for discrete X and similarly for continuous X:

$$P(|T - \theta| > \varepsilon) \le \frac{E((T-\theta)^2)}{\varepsilon^2} = \frac{MSE(T)}{\varepsilon^2}$$

Proving Markov's inequality (see above) for discrete variables X:

$$E(X^2) = \sum_x x^2 P(X = x) = \sum_{|x| \le \varepsilon} x^2 P(X = x) + \sum_{|x| > \varepsilon} x^2 P(X = x)$$

$$\ge \sum_{|x| > \varepsilon} x^2 P(X = x) \ge \sum_{|x| > \varepsilon} \varepsilon^2 P(X = x) = \varepsilon^2 \sum_{|x| > \varepsilon} P(X = x) = \varepsilon^2 P(|X| > \varepsilon)$$

Hence $P(|X| > \varepsilon) \leq \frac{E(X^2)}{\varepsilon^2}$ (and for continuous X we have a similar proof).