Welcome to your second homework. Here we begin our discussion of Point Estimation: Unbiasedness, Efficiency, Sufficiency.

As usual, try to find mistakes (and get extra points) in my solutions. Typically they are silly arithmetic mistakes (not methodological ones). They allow me to check that you did your HW on your own. Please do not e-mail me about your findings — just mention them on the first page of your solution and count extra points.

Now let us look at your problems.

1. Problem 10.2. We have a parametric family of probability spaces \((S, \mathcal{F}, P_\theta, \theta \in \Omega)\). It is assumed that an underlying parameter \(\theta\) is fixed but unknown to the statistician. It is given that two estimators, based on available observations from a space corresponding to an underlying \(\theta\), satisfy \(E_\theta(\hat{\Theta}_1) = \theta\) and \(E_\theta(\hat{\Theta}_2) = \theta\) for all considered \(\theta \in \Omega\). As a result, these two estimators are called unbiased.

Now a statistician suggests to consider a new estimator (a function of observations)

\[
\hat{\Theta}_3 = k_1 \hat{\Theta}_1 + k_2 \hat{\Theta}_2.
\]

Note that this new estimator is a linear combination of the former two. The statistician wants this new estimator to be unbiased as well. To this end, we need \(E_\theta(\hat{\Theta}_3) = \theta\) for all \(\theta \in \Omega\). Write, using the property that the expectation of a sum is equal to the sum of expectations and that the expectation of a factor times RV is the factor times the expectation of the RV,

\[
E_\theta(\hat{\Theta}_3) = k_1 E_\theta(\hat{\Theta}_1) + k_2 E_\theta(\hat{\Theta}_2) = (k_1 + k_2)\theta.
\]

This implies that the linear combination of the two unbiased estimators is again an unbiased estimator iff \(k_1 + k_2 = 1\). This is the answer.

2. Problem 10.5. Here we have a sample of size \(n\) from a population with the known mean \(\mu\) and the finite variance \(\sigma^2\). Then we can write:

\[
E\{n^{-1} \sum_{i=1}^{n} (X_i - \mu)^2\} = n^{-1} \sum_{i=1}^{n} E(X_i - \mu)^2 = n^{-1} \sum_{i=1}^{n} \sigma^2 = \sigma^2.
\]

Note that the last line holds for any underlying \(\mu\) and \(\sigma\), and thus the estimator is unbiased.

3. Problem 10.6. Here we again have a sample of size \(n\) from a population with the known mean \(\mu\) and the finite variance \(\sigma^2\). Then we can write:

\[
E\bar{X}^2 = E(\bar{X} - \mu + \mu)^2 = E(\bar{X} - \mu)^2 + \mu^2.
\]

As we know (or check this by a direct calculation)

\[
Var(\bar{X}) = E(\bar{X} - \mu)^2 = \sigma^2/n.
\]

Thus we get

\[
E(\bar{X}^2) = \mu^2 + \sigma^2/n,
\]
which yields that $\bar{X}^2$ (the squared sample mean, and recall that $\bar{X}$ is always an unbiased estimator of $\mu$) is an asymptotically unbiased estimator of $\mu^2$.

4. Problem 10.13. Let $E_{\theta}(\hat{\Theta}) = \theta$, $\theta \in \Omega$; this means that the estimator is unbiased. It is given that $Var(\hat{\Theta}) \neq 0$. Then we can write
\[
E_{\theta}(\hat{\Theta}^2) = E_{\theta}(\hat{\Theta} - \theta + \theta)^2 = E_{\theta}(\hat{\Theta} - \theta)^2 + \theta^2
\]
\[
= Var(\hat{\Theta}) + \theta^2.
\]
Remark: Please recall that what is written above is the familiar formula stating that the variance of a RV is equal to the expectation of the squared RV minus squared mean of the RV, and here due to unbiasedness the mean of the RV (the estimator) is equal to the parameter.

We conclude that $\hat{\Theta}^2$ is not an unbiased estimator of $\theta^2$.

5. Problem 10.15. Let $X_1, \ldots, X_n$ be iid Poisson($\lambda$). Recall that $E_{\lambda}(X_i) = \lambda$ and $Var_{\lambda}(X_i) = \lambda$. Also, as usual, $E_{\lambda}(\bar{X}) = \lambda$ for any $\lambda > 0$. This yields that $\bar{X}$ is an unbiased estimator of the parameter $\lambda$.

To evaluate its efficiency (the minimum variance over all unbiased estimators), we shall use Cramer-Rao inequality. To use it, we need to calculate variance of the estimator and compare it with the reciprocal of Fisher information for the sample: if the two coincide then the estimator is efficient.

Well, on one hand we have
\[
Var_{\lambda}(\bar{X}) = Var_{\lambda}(X_1)/n = \lambda/n.
\]
On the other hand, we can calculate Fisher information. Recall that Poisson pmf is $f_{\lambda}(x) = e^{-\lambda} \lambda^x / x!$, $x = 0, 1, \ldots$, and
\[
I_X(\lambda) = E_{\lambda}(\partial \ln(f_{\lambda}(X))/\partial \lambda)^2
\]
\[
= E_{\lambda}[\partial(-\lambda + X \ln(\lambda) - \ln(X!))/\partial \lambda]^2
\]
\[
= E_{\lambda}[-1 + X/\lambda]^2 = E_{\lambda}[1 - 2X/\lambda + X^2/\lambda^2] =
\]
\[
= 1 - 2\lambda/\lambda + (\lambda + \lambda^2)/\lambda^2 = 1/\lambda.
\]

Here the sample is a sample of iid Poisson RVs, so $I_{(X_1, \ldots, X_n)}(\lambda) = nI_X(\lambda)$. Thus we conclude that here we have equality in the Cramer-Rao inequality:
\[
Var_{\lambda}(\bar{X}) = \frac{1}{nI_X(\lambda)}.
\]
This establishes efficiency of the sample mean estimate among all unbiased estimators of $\lambda$.

6. Problem 10.16. Because $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are independent, and using additional information that these estimators are unbiased estimators of the parameter $\theta$, and $Var_{\theta}(\hat{\Theta}_1) = 3Var_{\theta}(\hat{\Theta}_2)$, we can write for $\hat{\Theta}_3 := a_1\hat{\Theta}_1 + a_2\hat{\Theta}_2$:
\[
E_{\theta}(\hat{\Theta}_3) = (a_1 + a_2)\theta
\]
and 

\[ \text{Var}(\hat{\Theta}_3) = a_1^2 \text{Var}_{\theta}(\hat{\Theta}_1) + a_2^2 \text{Var}_{\theta}(\hat{\Theta}_2) \]

\[ = (3a_1^2 + a_2^2) \text{Var}(\hat{\Theta}_2). \]

Now we are using those results in turn. First, for \( \hat{\Theta}_3 \) to be an unbiased estimator we must have \( a_1 + a_2 = 1. \) For its variance this implies that 

\[ 3a_1^2 + a_2^2 = 3(1 - 2a_2 + a_2^2) + a_2^2 = 3 - 6a_2 + 4a_2^2. \]

To minimize the variance, we need to minimize in \( a_2 \) the above-written expression. That parabola takes on its minimal value (take derivative, set it to zero, solve the equation, and check that this is the point of minimum) on \( a_2^* = 3/4. \) Then \( a_1^* = 1/4. \)

Answer: Choose \((a_1, a_2) = (1/4, 3/4).\)

7. Problem 10.21. Let \( \bar{X}_1 \sim N(\mu, \sigma_1^2/n) \) and \( \bar{X}_2 \sim N(\mu, \sigma_2^2/n). \) It is also given that these two sample means are independent. [Note that the latter information simplifies calculation of the variance of any linear combination of the two statistics.]

Then we study a linear combination \( \delta := w\bar{X}_1 + (1-w)\bar{X}_2 \) with \( w \in (0,1). \)

(a) Let us establish unbiasedness of \( \delta. \) Write,

\[ E_{\mu}(\delta) = wE_{\mu}(\bar{X}_1) + (1-w)E_{\mu}(\bar{X}_2) = (w + (1-w))\mu = \mu. \]

What was wished to show.

(b) Let us explore the variance of the linear combination. Write,

\[ \text{Var}(\delta) = w^2\text{Var}(\bar{X}_1) + (1-w)^2\text{Var}(\bar{X}_2) \]

\[ = w^2\sigma_1^2/n + (1-w)^2\sigma_2^2/n = n^{-1}[w^2(\sigma_1^2 + \sigma_2^2) - 2w\sigma_2^2 + \sigma_2^2]. \]

The minimum of this variance in \( w \) is attained at \( w^* = \sigma_2^2/(\sigma_1^2 + \sigma_2^2). \) [To see the latter take derivative, set it to zero and solve the equation.] Please note that the optimal \( w^* \) has sense, because the larger \( \sigma_2^2/\sigma_1^2 \) the larger our “trust” into \( \bar{X}_1. \)

8. Problem 10.33. Let \( Y \sim Unif(\alpha, \alpha + 1) \) and \( Y_{(1)} \) is the smallest observation among a sample \((Y_1, \ldots, Y_n)\) of size \( n. \) Then for any \( 0 < \epsilon < 1 \) we can write

\[ P(Y_{(1)} > \alpha + \epsilon) = P(Y_1 > \alpha + \epsilon, \ldots, Y_n > \alpha + \epsilon) \]

\[ = \prod_{i=1}^{n} P(Y_i > \alpha + \epsilon) = \prod_{i=1}^{n}(1-\epsilon) = (1-\epsilon)^n. \]

We established that

\[ P(|Y_{(1)} - \alpha| > \epsilon) = (1-\epsilon)^n \to 0 \quad \text{as} \quad n \to \infty. \]

Thus \( Y_{(1)} \) is a consistent estimator of \( \alpha. \)
9. Problem 10.36. Let \( X_1, \ldots, X_n \) be iid from \( Expon(\theta) \). Then, as we known from the theory of exponential distribution \( E_\theta(X) = E_\theta(X_1) = \theta \) and \( \text{Var}_\theta(X) = \text{Var}_\theta(X_1)/n = \theta^2/n \). Using these results, together with Chebyshev inequality, yields

\[
P(|\bar{X} - \theta| > \epsilon) \leq \frac{\text{Var}_\theta(\bar{X})}{\epsilon^2} = \frac{\theta^2}{\epsilon^2 n} \to 0 \quad \text{as} \quad n \to \infty.
\]

This proves that \( \bar{X} \) is a consistent estimator of \( \theta \).

10. Problem 10.42. Let \( X_1, \ldots, X_n \) be iid from \( Expon(\theta) \) distribution. The corresponding joint pdf is

\[
f_{\theta}(X_1, \ldots, X_n)(x_1, \ldots, x_n) = \prod_{i=1}^{n} \left( \frac{1}{\theta} \right) e^{-x_i/\theta} = \theta^{-n} e^{-\bar{X}/\theta} n! x_1!(n_1 - x_1)! x_2!(n_2 - x_2)! \theta^{x_1+x_2}(1 - \theta)^{n_1+n_2-(x_1+x_2)}
\]

This factorization in writing the joint pdf, according to the Factorization Theorem, yields sufficiency of the statistic \( \bar{X} \).

11. Problem 10.43. Let \( X_1 \) and \( X_2 \) be independent and \( X_1 \sim Bin(\theta, n_1) \) and \( X_1 \sim Bin(\theta, n_1) \). Then

\[
f_{\theta}(X_1, X_2)(x_1, x_2) = f_{\theta}^{X_1}(x_1) f_{\theta}^{X_2}(x_2) \quad \text{[this is due to independence]}
\]

\[
= \frac{n_1!}{x_1!(n_1 - x_1)!} \frac{n_2!}{x_2!(n_2 - x_2)!} \theta^{x_1}(1 - \theta)^{n_1-x_1} \theta^{x_2}(1 - \theta)^{n-x_2}
\]

\[
= \frac{n_1! n_2!}{x_1!(n_1 - x_1)! x_2!(n_2 - x_2)!} \theta^{x_1+x_2}(1 - \theta)^{n_1+n_2-(x_1+x_2)}
\]

This factorization, together with the Factorization Theorem, implies that \( X_1 + X_2 \) is a sufficient statistic. Plainly a one-to-one transformation of this statistic into a statistic \( (X_1 + X_2)/(n_1 + n_2) \) is also a sufficient statistic. Note that the latter is also an unbiased estimator.

12. Problem 10.48. Let \( X_1, \ldots, X_n \) be iid from \( Geom(\theta) \) distribution. Recall that \( X_i \) is the number of failures until first success (sometimes we use the total number of trials until the first success — the difference is just 1 trial between the two approaches). Then \( Y := \sum_{i=1}^{n} X_i \) is the number of failures until \( n \)th success, and it has a negative binomial distribution. Also recall that \( f_{\theta}^{X_1}(x) = (1 - \theta)^x \theta \).

Now we can write,

\[
f_{\theta}(X_1, \ldots, X_n)(x_1, \ldots, x_n) = (1 - \theta)^{\sum_{i=1}^{n} x_i} \theta^n.
\]

This and the Factorization Theorem yield that \( Y \) is the sufficient statistic.