1. (a) The set of possible values of \( S \) is \( \{n, n+1, \ldots, 2n\} \). For \( k \) in this set, \( S = k \) if exactly \( k-n \) of the \( n \) \( X_i \)'s are equal to two. Hence

\[
p_S(k) = p_B(k - n) = \binom{n}{k-n} p^{k-n} (1-p)^{n-(k-n)} = \binom{n}{k-n} p^{k-n} (1-p)^{2n-k}
\]

for \( n \leq k \leq 2n \).

ALTERNATIVELY, we see that \( Y_i = X_i - 1 \) is a Bernoulli random variable with parameter \( p \), and \( S = n + B \), where \( B = Y_1 + \cdots + Y_n \), so \( B \) has the binomial distribution with parameters \( n \) and \( p \). Hence, for \( n \leq k \leq 2n \), \( p_S(k) = p_B(k - n) = \binom{n}{k-n} p^{k-n} (1-p)^{n-(k-n)} \).

(b) \( E[X_1] = p \cdot 2 + (1-p) \cdot 1 = 1 + p \). Hence, \( E[S] = nE[X_1] = n(1+p) \).

ALTERNATIVELY, \( E[S] = E[B + n] = E[B] + n = np + n = n(1+p) \).

(c) \( \text{Var}(X_1) = E[X_1^2] - E[X_1]^2 = 4p + (1-p) - (1+p)^2 = p(1-p) \). So, since the \( X_i \)'s are independent (hence uncorrelated), \( \text{Var}(S) = n\text{Var}(X_1) = np(1-p) \).

ALTERNATIVELY, \( \text{Var}(S) = \text{Var}(B + n) = \text{Var}(B) = np(1-p) \).

2. (a)

\[
F_X(c) = \begin{cases} 
0, & c < 0 \\
\int_0^c e^{-5u} du = -e^{-5u} | \bigg |_0^c = 1 - e^{-5c}, & c > 0 
\end{cases}
\]

(b) \( P\{X > 1\} = 1 - P\{X \leq 1\} = 1 - F_X(1) = e^{-5} \)

(c) \( P\{X > 1 \mid X \leq 2\} = \frac{P\{X > 1, X \leq 2\}}{P\{X \leq 2\}} = \frac{F_X(2) - F_X(1)}{F_X(2)} = e^{-5} - e^{-10} \)

3. (a) No. For example, \( P\{T > 1\} = 1 - F_T(1) = e^{-1} \), whereas

\[
P(T > 2 \mid T > 1) = \frac{P(T > 2, T > 1)}{P(T > 1)} = \frac{P(T > 2)}{P(T > 1)} = \frac{1 - F_T(2)}{1 - F_T(1)} = \frac{e^{-4}}{e^{-1}} = e^{-3}
\]

(b) For \( x > 0 \)

\[F_X(x) = P\{X \leq x\} = P\{T \leq \sqrt{x}\} = 1 - e^{-x}\]

and \( f_X(x) = e^{-x} \), i.e., \( x \) has an exponential distribution.

(c) Yes, as we proved in class, the exponential distribution is memoryless. For any \( s, t > 0 \),

\[P(T > t + s \mid T > t) = \frac{P(T > t + s, T > t)}{P(T > t)} = \frac{P(T > t + s)}{P(T > t)} = \frac{e^{-(t+s)}}{e^{-t}} = e^{-s} = P\{T > s\} \]

4. (a) The time between emails is an exponential random variable with mean \( 1/\lambda = 1/5 \), so you expect to wait 2/5 hours.

(b) The number of emails in a 50 minute period is a Poisson random variable with mean \( \bar{\lambda} = \lambda(50/60) = 25/6 \), so you expect 25/6 emails.
(c) The number of emails in a 50 minute period is a Poisson random variable with mean
\( \lambda = \lambda \frac{50}{60} = 25/6 \), so
\[ P\{5 \text{ emails in 50 min} \} = \frac{e^{-\lambda} \lambda^5}{5!} = \frac{e^{-25/6} (25/6)^5}{5!}. \]

(d) Let \( R_1 \) be the arrival time of the first real news email, and let \( F_1 \) be the arrival time of
the first fake email.

\[
P\{\text{first email you receive is fake} \} = P\{F_1 < R_1 \} = \int_{0}^{\infty} P\{F_1 < R_1 | R_1 = u\} f_R(u) du
\]

= \[
\int_{0}^{\infty} (1 - e^{-10u}) 5e^{-5u} du = 5 \left[ \frac{e^{-5u}}{-5} - \frac{e^{-15u}}{-15} \right]_{0}^{\infty} = 5 \left( \frac{1}{5} - \frac{1}{15} \right) = \frac{10}{15} = \frac{2}{3}
\]

5. (a) The Gaussian density, 
\[ f_X(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(u-\mu)^2}{2\sigma^2} \right) \]
is maximized at \( u = \mu \). The maximum value is
\[ f_X(\mu) = \frac{1}{\sqrt{2\pi}\sigma} = \frac{1}{\sqrt{32\pi}} \approx 0.1. \]

(b) \( f_X(u) = (0.5)f_X(\mu) \) if the exponential term in the pdf at \( u \) is equal to 0.5. That is, if
\( \frac{(u-\mu)^2}{2\sigma^2} = \ln(2) \) or \( (u-\mu)^2 = 2(\ln 2)\sigma^2 \) or \( u = \mu \pm \sigma \sqrt{2 \ln 2} = 16 \pm 4\sqrt{2 \ln 2} \approx 16 \pm 4.7. \)

(c) A matlab plot is shown. Important features are that the pdf is centered at 20, the maximum is around 0.1, and the shape is a bell curve with halfwidth around 5.

![Gaussian density](image.png)

(d) \( P\{|X| \geq 30\} = P\{X \geq 30\} + P\{X \leq -30\} = P\{X = \frac{X - 20}{4} \geq \frac{30 - 20}{4}\} + P\{X = \frac{X - 20}{4} \leq \frac{-30 - 20}{4}\} =
Q(2.5) + 1 - Q(-12.5) = Q(2.5) + Q(12.5) \approx Q(2.5) \approx 0.0062. \)

6. By definition, \( \hat{\theta}_{ML}(10) \) is the value of \( \theta \) that maximizes the likelihood of \( X = 10 \). The likelihood of \( X = 10 \) is \( f_\theta(10) = \left(\frac{10}{\theta}\right) e^{-\frac{100}{\theta}} \), and the log likelihood is \( \ln(10) - \ln(\theta) - \frac{50}{\theta} \). Differentiation with respect to \( \theta \) yields \( \frac{d\ln f_\theta(10)}{d\theta} = -\frac{1}{\theta} + \frac{50}{\theta^2} \). This derivative is zero for \( \theta = 50 \), and it is positive for \( \theta < 50 \) and negative for \( \theta > 50 \). Hence, \( \hat{\theta}_{ML}(10) = 50 \).

(Note: If \( \theta \) is replaced by \( \sigma^2 \), then \( f \) is the Rayleigh pdf with parameter \( \sigma^2 \). The same reasoning as above shows that, in general, for observation \( X = u, \hat{\sigma}_{ML} = \left(\frac{\sigma^2}{\hat{\sigma}_{ML}}\right)_{ML} = \frac{u^2}{2} \).)

7. (a) The range of possible values of \( X \) is the interval \([0,1] \). By inspection (or using the
definition of conditional density), if \( 0 \leq u \leq 1 \), we see that given \( X = u \), the conditional
distribution of \( Y \) is the uniform distribution over the interval \([0,u] \). Therefore, \( E[Y^2 | X = u] = \int_0^u \frac{1}{2} u^2 dv = \frac{u^2}{3} \).
(b) By the same observation used in part (a), \( E[Y|X = u] = \frac{u}{2} \). Since this is a linear function of \( u \), it is also equal to \( \hat{E}[Y|X = u] \). That is, \( \hat{E}[Y|X = u] = \frac{u}{2} \). (This result can be worked out using the formula for \( \hat{E}[Y|X = u] \) as well.)

8. (a) First note that \( \text{Cov}(X, Y) = \sigma_X \sigma_Y \rho = \frac{1}{4} \). Therefore
\[
\text{Cov}(X, W) = \text{Cov}(X, X + \alpha Y + \beta) = \text{Var}(X) + \alpha \text{Cov}(X, Y) = 1 + \frac{\alpha}{4},
\]
and \( \text{Cov}(X, W) = 0 \) if \( \alpha = -4 \), no matter what value \( \beta \) takes. Since \( X \) and \( W \) are jointly Gaussian, \( \alpha = -4 \) (and \( \beta \) arbitrary) also makes them independent.

(b) \( E[Z] = E[4X + 2Y + 2] = 4E[X] + 2E[Y] + 3 = 7 \), and \( \text{Var}(Z) = 16 \text{Var}(X) + 4 \text{Var}(Y) + 16 \text{Cov}(X, Y) = 16 + 16 + 4 = 36 \)

(c) Since \( Y \) and \( Z \) are jointly Gaussian
\[
E[Y|Z = 11] = L^*(11) = E[Y] + \frac{\text{Cov}(Y, Z)}{\text{Var}(Z)} (11 - E[Z])
\]
Now
\[
\text{Cov}(Y, Z) = \text{Cov}(Y, 4X + 2Y + 2) = 4 \text{Cov}(Y, X) + 2 \text{Var}(Y) = 1 + 8 = 9
\]
and therefore
\[
E[Y|Z = 11] = L^*(11) = 0 + \frac{9}{36} (11 - 7) = 1
\]

(d) Given, \( Z = 11 \), \( Y \) is Gaussian with mean \( E[Y|Z = 11] = 1 \) as calculated above, and variance
\[
\sigma^2_e = \text{MSE of } L^*(11) = \text{Var}(Y) - \frac{(\text{Cov}(Y, Z))^2}{\text{Var}(Z)} = 4 - \frac{9^2}{36} = 4 - \frac{9}{4} = \frac{7}{4}
\]
Therefore
\[
E[Y^2|Z = 11] = 1 + \frac{7}{4} = \frac{11}{4}
\]

9. (a)
\[
f_Y(v) = \begin{cases} 
  \int_{-\infty}^{v} \frac{1}{2} e^{-u} \, du = -\frac{1}{2} e^{-u} \big|_{-\infty}^{v} = \frac{1}{2} e^v, & v < 0 \\
  \int_{v}^{\infty} \frac{1}{2} e^{-u} \, du = \frac{1}{2} e^{-v}, & v > 0
\end{cases} = \frac{1}{2} e^{-|v|}
\]

(b) \( E\left[\frac{1}{X}\right] = \int_{0}^{\infty} \int_{-\infty}^{1-\frac{u}{2}} e^{-u} \, du \, dv = \int_{0}^{\infty} e^{-v} \, dv = 1 \)

10. (a)
\[
f_X(u) = \begin{cases} 
  \int_{u}^{1} 2 \, dv = 2(1 - u), & 0 < u < 1 \\
  0, & \text{elsewhere}
\end{cases}
\]

\[\implies f_{Y|X}(v|X = u) = \frac{1}{1 - u}, \quad 0 < u < 1, \ u \leq v < 1\]
That is, given \( X = u \), \( Y \) is uniformly distributed over the interval [\( u, 1 \)], as can also be seen by inspection. Therefore,

\[ \text{and } \ g^*(u) = E[Y|X = u] = \int_{u}^{1} \frac{v}{1-u} \, dv = \frac{1 + u}{2}, \quad 0 < u < 1 \]
(b) 

\[ E[Y^2|X = u] = \int_u^1 \frac{v^2}{1-u} \, dv = \frac{1+u+u^2}{3} \]

\[ \Rightarrow MMSE = Var(Y|X = u) = E[Y^2|X = u] - (E[Y|X = u])^2 = \frac{(1-u)^2}{12} \]

(c) Since the unconstrained estimator is linear in \( u \), the best linear estimator is:

\[ E[Y|X = u] = \frac{1+u}{2} \]
11. (a) False, False, True  
(b) True, False, False  
(c) True, True  
(d) True, False