MT2.1 (20 Points) Consider a random variable X whose transform (i.e., moment generating function) is $M_X(s)$. For each candidate transform expression in parts (a) to (d), select the strongest correct statement from the choices below:

(I) The expression *is* a valid transform associated with a random variable.

(II) The expression *can* be a valid transform associated with a random variable, but more information is needed to reach a definitive conclusion.

(III) The expression *cannot* be a valid transform associated with a random variable.

If you choose option (I), express the new random variable (whose transform is given) in terms of X. If you choose option (II), find one random variable having the given transform expression, and establish its relationship with X. Whatever your choice, provide a succinct, but clear and convincing, explanation.

(a) (5 Points) $M_V(s) = 3\,M_X(2s)$.

(III). Any transform (moment generating function) $M_V(s)$ must evaluate to unity at $s = 0$: in particular,

$$M_V(0) = \int_{-\infty}^{+\infty} e^{0\cdot v} f_V(v)\,dv = \int_{-\infty}^{+\infty} f_V(v)\,dv = 1.$$

Since $M_V(0) = 3M_X(0) = 3 \neq 1$, it cannot be a valid transform..

(b) (5 Points) $M_W(s) = M_X(s)M_X(-s)$.

(I). Let Y be a random variable independent of X, but whose density $f_Y(y) = f_X(-x)$. We now show that $M_Y(s) = M_X(-s)$:

$$M_Y(s) = \mathbf{E}\left[e^{sY}\right] = \int_{-\infty}^{+\infty} e^{sy} f_Y(y)\,dy = \int_{-\infty}^{+\infty} e^{sy} f_X(-y)\,dy$$

Now let $\tau = -y$, so $dy = -d\tau$:

$$M_Y(s) = -\int_{+\infty}^{-\infty} e^{-s\tau} f_X(\tau)\,d\tau = \int_{-\infty}^{+\infty} e^{-s\tau} f_X(\tau)\,d\tau = M_X(-s).$$

The variable W is then given by $W = Y + X$, which has the transform $M_W(s) = M_X(s)M_X(-s)$. A similar reasoning works if X is a discrete random variable having PMF $p_X(x)$.
Please note that Y has the same density as random variable $-X$, but this does *not* mean $Y = -X$.

(c) (5 Points) $M_Y(s) = M_Q(s)M_X(s)$, where $M_Q(s) = \exp\left[2(e^s - 1)\right]$.

(I). Let Q be a Poisson random variable with parameter $\lambda = 2$. This means

$$p_Q(q) = \frac{2^q e^{-2}}{q!} u(q),$$

where q is an integer and u is the discrete unit-step function. Assume also that Q is independent of X. Then $Y = Q + X$ has the desired transform $M_Y(s) = M_Q(s)M_X(s)$.

(d) (5 Points) $M_Z(s) = M_R(s)M_X(s)$, where $M_R(s) = \frac{1}{6}e^{-3s} + \frac{1}{2}e^{-s} + \frac{1}{3}e^{5s}$.

(I). Let R be a discrete random variable independent of X, and with probability mass function

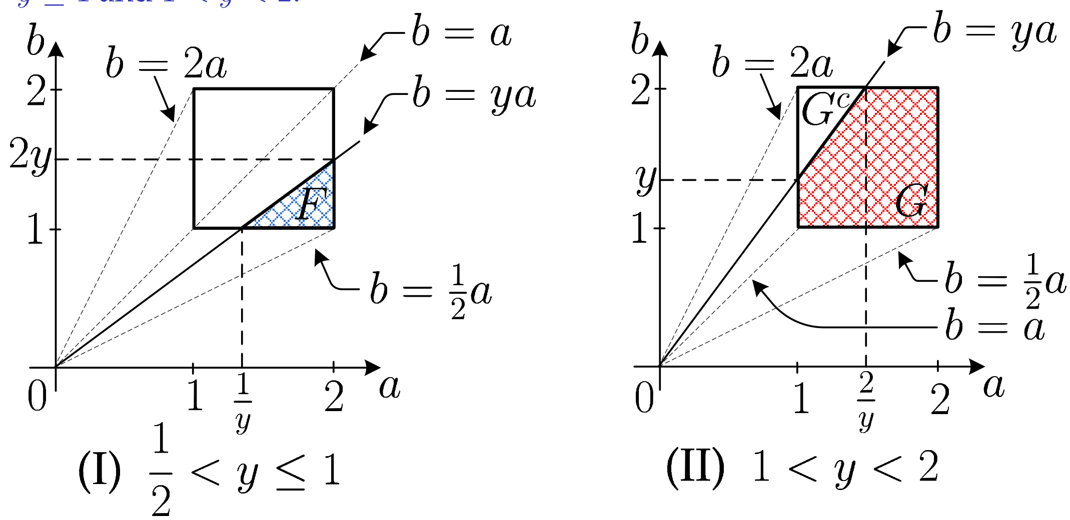
$$p_R(r) = \frac{1}{6}\delta(r+3) + \frac{1}{2}\delta(r+1) + \frac{1}{3}\delta(r-5).$$

Then $Z = R + X$ has the desired transform $M_Z(s) = M_R(s)M_X(s)$.

MT2.2 (30 Points) Consider a random quadratic polynomial $Q(x) = Ax^2+Bx+C$, where A, B, and C are mutually independent random variables uniformly distributed over the interval $[1, 2]$. Let $\widehat{X}$ denote the value of x corresponding to the extremum (global minimum or maximum) of the polynomial Q.

(a) (15 Points) Determine, and provide a well-labeled plot of, $f_{\widehat{X}}(\widehat{x})$, the PDF of $\widehat{X}$.
The extremum point is the solution to the equation $Q'(x) = 2Ax+B = 0$, which is $\widehat{X} = -\frac{B}{2A}$. Instead of looking at $\widehat{X}$ directly, though, let's look at its more modest, less cluttered cousin Y defined as $Y = B/A$. We'll first determine the PDF $f_Y(y)$ from the CDF $F_Y(y)$. Since $\widehat{X}$ and Y are related linearly—in particular, $\widehat{X} = -Y/2$—it's then easy to determine $f_{\widehat{X}}(\widehat{x})$ from $f_Y(y)$.

Clearly, $Y : \Omega \mapsto [1/2, 2]$, so $F_Y(y)$ is 0 for $y \leq 1/2$ and it's 1 for $y \geq 2$. To determine $F_Y(y)$ in the interval $(1/2, 2)$ we must consider two qualitatively different intervals, as depicted by Figures (I) and (II). These intervals are $1/2 < y \leq 1$ and $1 < y < 2$.



We know that $F_{A,B}(a, b) = 1$ in the square region extending from 1 to 2 along each axis, and is zero elsewhere. For $1/2 < y \leq 1$, the CDF $F_Y(y)$ is determined simply by looking at the probability of Event F, which is the area of its corresponding triangle (recall that the joint PDF $F_{A,B}(a, b) = 1$ there). That is,

$$F_Y(y) = \frac{1}{2}(2y-1)\left(2-\frac{1}{y}\right) = 2y-2+\frac{1}{2y} \quad \text{for} \quad 1/2 < y \leq 1.$$

Of course, we can also obtain this by integration, but why bother?

$$F_Y(y) = \int_{1/y}^{2}\int_{1}^{ya} db\, da = \int_{1/y}^{2}(ya-1)\, da = \left[y\frac{a^2}{2}-a\right]_{1/y}^{2} = \cdots$$

We now consider the interval $1 < y < 2$. The CDF $F_Y(y)$ is then the probability corresponding to the shaded region marked G in Figure (II). However, it's easier to determine the probability of the complementary event G^c and subtract it from 1, so that's what we'll do below:

$$F_Y(y) = 1 - \frac{1}{2}\left(\frac{2}{y} - 1\right)(2-y) = 3 - \frac{2}{y} - \frac{y}{2} \quad \text{for} \quad 1 < y < 2.$$

A more tedious method, which uses integration, goes as follows:

$$F_Y(y) = \int_1^{2/y}\int_1^{ya} db\,da + \int_{2/y}^{2}\int_1^{2} db\,da = \int_1^{2/y}(ya-1)\,da + \int_{2/y}^{2} da = \cdots$$

Now we have the CDF for Y:

$$F_Y(y) = \begin{cases} 0 & \text{if } y \le 1/2 \\ 2y - 2 + \dfrac{1}{2y} & \text{if } 1/2 < y \le 1 \\ 3 - \dfrac{2}{y} - \dfrac{y}{2} & \text{if } 1 < y < 2 \\ 1 & \text{if } 2 \le y. \end{cases}$$

To obtain the PDF $f_Y(y)$, we simply differentiate the CDF $F_Y(y)$:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \begin{cases} 2 - \dfrac{1}{2y^2} & \text{if } 1/2 < y \le 1 \\ \dfrac{2}{y^2} - \dfrac{1}{2} & \text{if } 1 < y < 2 \\ 0 & \text{elsewhere.} \end{cases}$$

In shorthand, $f_Y(y) = \left(2 - \frac{1}{2y^2}\right)\mathbf{1}_{(1/2,1)}(y) + \left(\frac{2}{y^2} - \frac{1}{2}\right)\mathbf{1}_{(1,2)}(y)$, where the indicator function $\mathbf{1}_{(\alpha,\beta)}(y) = 1$ if $y \in (\alpha, \beta)$ and it's zero elsewhere. Now, $\widehat{X} = -Y/2$. We know that if $\widehat{X} = \lambda Y + \mu$, for constants λ and μ, where $\lambda \neq 0$, then

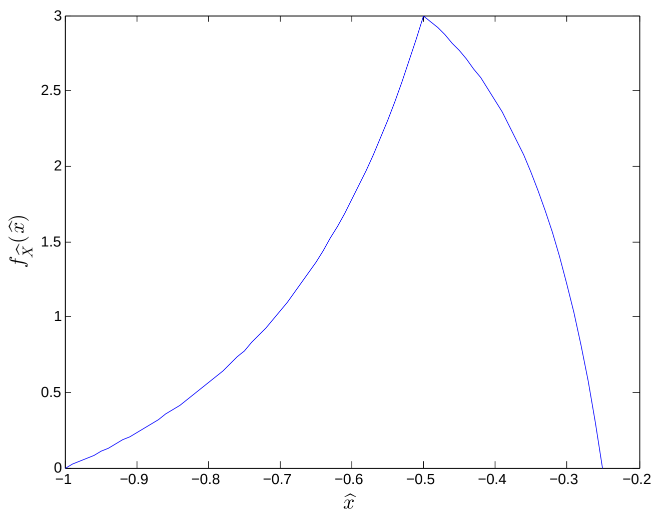
$$f_{\widehat{X}}(\widehat{x}) = \frac{1}{|\lambda|} f_Y\left(\frac{\widehat{x} - \mu}{\lambda}\right).$$

In our case $\lambda = -1/2$ and $\mu = 0$. Therefore, $f_{\widehat{X}}(\widehat{x}) = \frac{1}{|-1/2|} f_Y(-2\widehat{x}) = 2 f_Y(-2\widehat{x})$. The interval $1/2 < y \le 1$ corresponds to $-1/2 \le \widehat{x} < -1/4$, and the interval

$1 < y < 2$ corresponds to $-1 < \widehat{x} < -1/2$. The PDF for the extremum point is then given by

$$f_{\widehat{X}}(\widehat{x}) = \begin{cases} \dfrac{1}{\widehat{x}^2} - 1 & \text{if} \quad -1 < \widehat{x} < -1/2 \\ 4 - \dfrac{1}{4\widehat{x}^2} & \text{if} \quad -1/2 \leq \widehat{x} < -1/4 \\ 0 & \text{elsewhere.} \end{cases}$$

In shorthand $f_{\widehat{X}}(\widehat{x}) = \left(\frac{1}{\widehat{x}^2} - 1\right) \mathbf{1}_{(-1,-1/2)}(\widehat{x}) + \left(4 - \frac{1}{4\widehat{x}^2}\right) \mathbf{1}_{[-1/2,-1/4)}(\widehat{x})$.



(b) (10 Points) Determine $\mathbf{E}[\widehat{X}]$.

$$\mathbf{E}[\widehat{X}] = \mathbf{E}\left[-\frac{B}{2A}\right] = -\int_1^2 \int_1^2 \frac{b}{2a} db\, da$$
$$= -\int_1^2 \frac{3}{4a} da = -\frac{3}{4} \ln 2.$$

(c) (5 Points) Determine $\operatorname{cov}(\widehat{X}, A)$. Explain why your answer makes sense.

$$\begin{aligned} \operatorname{cov}(\widehat{X}, A) &= \mathbf{E}[\widehat{X}A] - \mathbf{E}[\widehat{X}]\,\mathbf{E}[A] \\ &= -\mathbf{E}\left[\frac{B}{2}\right] - \mathbf{E}[\widehat{X}]\,\mathbf{E}[A] \\ &= -\frac{3}{4} + \frac{9}{4} \ln 2. \end{aligned}$$

As expected, the covariance is positive.

MT2.3 (40 Points)

(a) (10 Points) Consider a random variable X that has a finite mean $\mathbf{E}[X]$, finite variance σ_X^2, and PDF $f_X(x)$. Suppose we want to estimate X with a constant parameter α. Then the quantity $X - \alpha$ denotes the *estimation error*. Show that the value of α that minimizes the *mean squared error* $\mathbf{E}[(X-\alpha)^2]$ is given by $\alpha = \mathbf{E}[X]$.

A similar result holds if we condition on an event A. In particular, the value of α that minimizes the mean squared error $\mathbf{E}\left[(X-\alpha)^2 | A\right]$ is given by $\alpha = \mathbf{E}[X|A]$. You need not show the result for the conditional case here; however, feel free to use it if you need to.

Method I: The *mean squared error* (MSE) is $\mathbf{E}[(X-\alpha)^2] = \mathbf{E}[X^2 - 2\alpha X + \alpha^2] = \mathbf{E}[X^2] - 2\alpha\mathbf{E}[X] + \alpha^2$. To determine the value of α that minimizes the MSE, differentiate the MSE with respect to α, set the result to zero, and solve for α. This leads to the equation $2\alpha - 2\mathbf{E}[X] = 0$, which yields $\alpha = \mathbf{E}[X]$.

Method II: Write the MSE as $\mathbf{E}\left[(X - \mathbf{E}[X] + \mathbf{E}[X] - \alpha)^2\right]$, which yields

$$
\begin{aligned}
\text{MSE} &= \mathbf{E}\left[(X-\mathbf{E}[X])^2 + 2(X-\mathbf{E}[X])(\mathbf{E}[X]-\alpha) + (\mathbf{E}[X]-\alpha)^2\right] \\
&= \mathbf{E}\left[(X-\mathbf{E}[X])^2\right] + 2(\mathbf{E}[X]-\alpha)\,\underbrace{\mathbf{E}[X-\mathbf{E}[X]]}_{=0} + \underbrace{\mathbf{E}\left[(\mathbf{E}[X]-\alpha)^2\right]}_{=(\mathbf{E}[X]-\alpha)^2} \\
&= \text{var}(X) + (\mathbf{E}[X]-\alpha)^2.
\end{aligned}
$$

To minimize the MSE, force the second term to zero by letting α equal $\mathbf{E}[X]$.

Method III: Let $\alpha = \mathbf{E}[X] + c$ for some constant c. Then

$$
\begin{aligned}
\text{MSE} &= \mathbf{E}\left[(X-\mathbf{E}[X]-c)^2\right] \\
&= \mathbf{E}\left[(X-\mathbf{E}[X])^2 - 2c(X-\mathbf{E}[X]) + c^2\right] \\
&= \mathbf{E}\left[(X-\mathbf{E}[X])^2\right] - 2c\,\underbrace{\mathbf{E}[X-\mathbf{E}[X]]}_{=0} + c^2 = \text{var}(X) + c^2.
\end{aligned}
$$

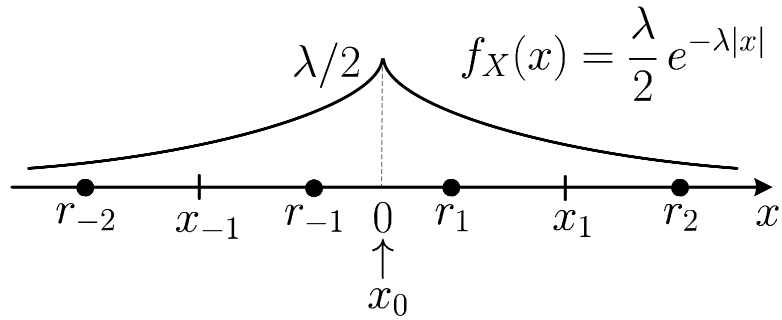
Let $c = 0$ to minimize the MSE.
Method IV: Write the MSE as

$$
\begin{aligned}
\mathbf{E}[(X-\alpha)^2] &= \mathbf{E}[X^2 - 2\alpha X + \alpha^2] = \mathbf{E}[X^2] - 2\alpha\mathbf{E}[X] + \alpha^2 \\
&= \underbrace{\mathbf{E}[X^2] - \mathbf{E}^2[X]}_{\text{var}(X)} + \underbrace{\mathbf{E}^2[X] - 2\alpha\mathbf{E}[X] + \alpha^2}_{=(\mathbf{E}[X]-\alpha)^2} = \text{var}(X) + (\mathbf{E}[X]-\alpha)^2.
\end{aligned}
$$

We can't do anything about the variance term. To minimize the MSE, α must equal $\mathbf{E}[X]$.

Please Note: Methods II, III, and IV are cosmetic variants of one another.

For the remainder of this problem, let X be a random variable whose PDF is the double-sided exponential shown below:



We want to encode X onto a two-bit binary number, an example of a discretization scheme known as *quantization*. We divide the x axis into a set of *quantization intervals* demarcated by the *decision boundaries* x_{-1}, x_0, and x_1, as shown in the figure above.

The quantized value Y is then defined as follows:

$$Y = \begin{cases} r_{-2} & \text{if } X < x_{-1} \\ r_{-1} & \text{if } x_{-1} \leq X < x_0 \\ r_1 & \text{if } x_0 \leq X < x_1 \\ r_2 & \text{if } x_1 \leq X. \end{cases}$$

Each of the values r_k is called a *reconstruction level*. We want to design our quantizer so that X is *equally likely* to be mapped to any of the four reconstruction levels. In other words, we want to design our decision boundaries x_{-1} and x_1 so that Y is equally likely to take on any of the values r_{-2}, r_{-1}, r_1, and r_2.

(b) (3 Points) Explain why x_{-1} must be equal to $-x_1$, and $r_{-k} = -r_k$, for $k = 1, 2$.

Due to symmetry of the PDF $f_X(x)$, we can perform the design using the portion of of it corresponding to $x \geq 0$. To satisfy the overall equiprobability design requirement, we then place the *decision boundaries* and the *reconstruction levels* symmetrically around $x = 0$. That is, we must have $x_{-1} = -x_1$ and $r_{-k} = -r_k$ for $k = 1, 2$.

This is sufficient to determine the exact locations of the decision boundaries x_1 and x_2. However, it's *not* sufficient to determine the exact locations of the reconstruction levels (beyond stipulating that they be symmetrically placed around zero).

(c) (10 Points) Determine the decision boundary x_1.

To design our quantizer so that X is *equally likely* to be mapped to any of the four reconstruction levels, the area under the PDF $f_X(x)$ between any consecutive decision boundaries must equal $\frac{1}{4}$. So, looking at the decision boundaries x_0 and x_1, we note that

$$\int_{x_0}^{x_1} f(x)dx = \int_0^{x_1} \frac{\lambda}{2}e^{-\lambda x}dx = \left[\frac{-e^{-\lambda x}}{2}\right]_0^{x_1} = \frac{1-e^{-\lambda x_1}}{2} = \frac{1}{4}$$

Solving for x_1 yields

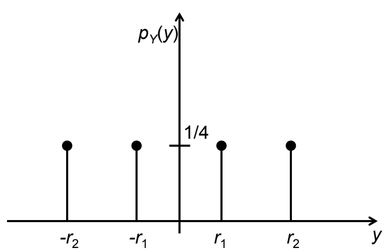
$$\boxed{x_1 = \frac{\ln(2)}{\lambda}}$$

We could have also looked at the decision boundaries x_1 and ∞. In particular,

$$\int_{x_1}^{\infty} f(x)dx = \int_{x_1}^{\infty} \frac{\lambda}{2}e^{-\lambda x}dx = \left[\frac{-e^{-\lambda x}}{2}\right]_{x_1}^{\infty} = \frac{e^{-\lambda x_1}}{2} = \frac{1}{4}$$

Solving for x_1 yields $x_1 = \ln(2)/\lambda$.

(d) (5 Points) Determine, and provide a well-labeled plot of, the PMF $p_Y(y)$. Also determine the mean $\mathbf{E}[Y]$, and the variance σ_Y^2. Your answers should be in terms of the reconstruction levels r_1 and r_2..



$$p_Y(y) = \begin{cases} \frac{1}{4} & \text{for } y = \pm r_1, \pm r_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{E}[Y] = \frac{r_{-2}+r_{-1}+r_1+r_2}{4} = \frac{-r_2-r_1+r_1+r_2}{4} = 0.$$

$$\begin{aligned} \sigma_Y^2 &= \mathbf{E}[(Y-\mathbf{E}[Y])^2] = \mathbf{E}[Y^2] - \mathbf{E}[Y]^2 = \mathbf{E}[Y^2] - 0 \\ &= \frac{r_{-2}^2+r_{-1}^2+r_1^2+r_2^2}{4} = \frac{(-r_2)^2+(-r_1)^2+r_1^2+r_2^2}{4} = \frac{r_1^2+r_2^2}{2} \end{aligned}$$

(e) (12 Points) Determine r_1 and r_2 to minimize the total distortion, defined as follows:

$$\begin{aligned}
\mathcal{D} &= \mathbf{E}\left[(X-Y)^2\right] = \sum_k \mathbf{E}\left[(X-Y)^2 | Y = r_k\right] \mathbf{P}(Y = r_k) \\
&= \sum_{k=\pm 1,2} \mathbf{E}\left[(X-Y)^2 | Y = r_k\right] \underbrace{\mathbf{P}(Y = r_k)}_{=1/4} \\
&= 2\sum_{k=1}^{2} \frac{1}{4}\mathbf{E}\left[(X-Y)^2 | Y = r_k\right] = \frac{1}{2}\sum_{k=1}^{2} \mathbf{E}\left[(X-Y)^2 | Y = r_k\right] \\
&= \frac{1}{2}\mathbf{E}\left[(X-Y)^2 | x_0 \le X < x_1\right] + \frac{1}{2}\mathbf{E}\left[(X-Y)^2 | x_1 \le X < x_2\right],
\end{aligned}$$

where $x_0 = 0$ and $x_2 = \infty$. To minimize $\mathcal{D}$ we must determine the values of r_1 and r_2 that minimize each of the two terms on the right-hand side, respectively. According to part (a), the α that minimizes $\mathbf{E}\left[(X-\alpha)^2 | A_k\right]$ is $\alpha = \mathbf{E}\left[X | A_k\right]$. The event A_k is defined in the following equivalent ways: $A_k = \{Y : Y = r_k\}$ or, equivalently, $A_k = \{X : x_{k-1} \le X < x_k\}$. Therefore,

$$\begin{aligned}
r_k = \mathbf{E}\left[X | A_k\right] &= \frac{\displaystyle\int_{x_{k-1}}^{x_k} x \frac{\lambda}{2} e^{-\lambda x} dx}{P(Y = r_k)} = \frac{\left[-x\,\frac{e^{-\lambda x}}{2}\right]_{x_{k-1}}^{x_k} + \frac{1}{2}\int_{x_{k-1}}^{x_k} e^{-\lambda x} dx}{\frac{1}{4}} \\
&= \left[-2xe^{-\lambda x}\right]_{x_{k-1}}^{x_k} - \left[\frac{2e^{-\lambda x}}{\lambda}\right]_{x_{k-1}}^{x_k} \\
&= 2x_{k-1}e^{-\lambda x_{k-1}} - 2x_k e^{-\lambda x_k} + \frac{2e^{-\lambda x_{k-1}}}{\lambda} - \frac{2e^{-\lambda x_k}}{\lambda}
\end{aligned}$$

So for $k = 1$, $x_0 = 0$, and $x_1 = \ln(2)/\lambda$,

$$\begin{aligned}
r_1 &= 2x_0 e^{-\lambda x_0} - 2x_1 e^{-\lambda x_1} + \frac{2e^{-\lambda x_0}}{\lambda} - \frac{2e^{-\lambda x_1}}{\lambda} = 0 - \frac{\ln(2)}{\lambda} + \frac{2}{\lambda} - \frac{1}{\lambda} \\
r_1 &= \frac{1 - \ln(2)}{\lambda}.
\end{aligned}$$

Similarly, for $k = 2$, $x_1 = \ln(2)/\lambda$, and $x_2 = \infty$.

$$\begin{aligned}
r_2 &= 2x_1 e^{-\lambda x_1} - 2x_2 e^{-\lambda x_2} + \frac{2e^{-\lambda x_1}}{\lambda} - \frac{2e^{-\lambda x_2}}{\lambda} = \frac{\ln(2)}{\lambda} - 0 + \frac{1}{\lambda} - 0 \\
r_2 &= \frac{1 + \ln(2)}{\lambda}.
\end{aligned}$$

It's not a coincidence that r_1 and r_2 are situated symmetrically around $1/\lambda$—the mean of a one-sided exponential with decay rate λ. After determining either of r_1, could we have determined r_2 without carrying any integration? The answer is yes—by exploiting the *Law of Total Expectation*. We encourage you to explore this further.

MT2.4 (15 Points) Consider a pair of IID Gaussian random variables X and Y each having a mean of zero and a variance equal to σ^2. Let Z be a third Gaussian random variable defined by $Z = X + Y$.

(a) (5 Points) Determine $f_{Z|X}(z|x)$.

Conditional on $X = x$, Z is simply Y shifted by x:

$$\begin{aligned} f_{Z|X}(z|x) &= f_{Y|X}(z-x|x) \\ &= f_Y(z-x) \\ &= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(z-x)^2}{2\sigma^2}}. \end{aligned}$$

(b) (10 Points) Determine $f_{X|Z}(x|z)$.

We know that Z, which is the sum of two independent Gaussian random variables, is Gaussian. We can obtain its mean and variance by summing those of X and Y. Hence we have

$$f_Z(z) = \frac{1}{2\sqrt{\pi}\,\sigma} e^{-\frac{z^2}{4\sigma^2}}.$$

We can now use Bayes's rule to compute

$$\begin{aligned} f_{X|Z}(x|z) &= \frac{f_{Z|X}(z|x) f_X(x)}{f_Z(z)} \\ &= \frac{\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(z-x)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}}}{\frac{1}{2\sqrt{\pi}\,\sigma} e^{-\frac{z^2}{4\sigma^2}}} \\ &= \frac{1}{\sqrt{\pi}\,\sigma} e^{-\frac{(2x-z)^2}{4\sigma^2}}. \end{aligned}$$

Note that this density corresponds to a Gaussian distribution of mean $z/2$ and variance $\sigma^2/2$.

LAST Name ____________________ FIRST Name ____________________

Recitation Time (Circle One): 10 11 12

Problem	Points	Your Score
Name	10	0
1	20	20
2	30	30
3	40	40
4	15	15
Total	**115**	**105**