

Question 1

(a) Conditional PDF of X given $Y = y$.

First marginalise over the support $0 \leq x \leq y$:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = 3 \int_0^y (1 + x - y) dx = 3 \left[(1 - y)x + \frac{x^2}{2} \right]_0^y = \frac{3}{2}y(2 - y), \quad 0 \leq y \leq 1.$$

The conditional-density formula gives, for $0 < y \leq 1$,

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \begin{cases} \frac{2(1 + x - y)}{y(2 - y)}, & 0 \leq x \leq y, \\ 0, & \text{otherwise.} \end{cases}$$

The restriction $y > 0$ avoids division by $f_Y(0) = 0$.

(b) MMSE estimate of X .

For $0 < y \leq 1$, the MMSE estimate is the conditional expected value:

$$\begin{aligned} \hat{x}_M(y) = E[X | Y = y] &= \int_{-\infty}^{\infty} x f_{X|Y}(x | y) dx \\ &= \frac{2}{y(2 - y)} \int_0^y ((1 - y)x + x^2) dx \\ &= \frac{2}{y(2 - y)} \left[\frac{(1 - y)x^2}{2} + \frac{x^3}{3} \right]_0^y = \boxed{\frac{y(3 - y)}{3(2 - y)}}, \quad 0 < y \leq 1. \end{aligned}$$

This is the range of observations for which the conditional-density calculation in (a) is defined; at $y = 0$, $f_Y(y) = 0$.

(c) Conditional PDF of Y given $X = x$.

Marginalise over $x \leq y \leq 1$, then apply the conditional-density formula:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = 3 \int_x^1 (1 + x - y) dy \\ &= 3 \left[(1 + x)y - \frac{y^2}{2} \right]_x^1 = \frac{3}{2}(1 - x^2), \quad 0 \leq x \leq 1, \\ f_{Y|X}(y | x) &= \frac{f_{X,Y}(x, y)}{f_X(x)} = \begin{cases} \frac{2(1 + x - y)}{1 - x^2}, & x \leq y \leq 1, \\ 0, & \text{otherwise,} \end{cases} \quad 0 \leq x < 1. \end{aligned}$$

The restriction $x < 1$ avoids division by $f_X(1) = 0$.

(d) MMSE estimate of Y .

$$\begin{aligned} \hat{y}_M(x) = E[Y | X = x] &= \int_{-\infty}^{\infty} y f_{Y|X}(y | x) dy \\ &= \frac{2}{1 - x^2} \int_x^1 ((1 + x)y - y^2) dy \\ &= \frac{2}{1 - x^2} \left[\frac{(1 + x)y^2}{2} - \frac{y^3}{3} \right]_x^1 \\ &= \frac{1 + 3x - 3x^2 - x^3}{3(1 - x^2)} = \boxed{\frac{x^2 + 4x + 1}{3(1 + x)}}, \quad 0 \leq x < 1. \end{aligned}$$

The last step cancels the common factor $1 - x$.

(e) Blind estimate and the benefit of observing Y .

$$\hat{x}_B = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \frac{3}{2} \int_0^1 (x - x^3) dx = \frac{3}{2} \left(\frac{1}{2} - \frac{1}{4} \right) = \boxed{\frac{3}{8}}.$$

Without observing Y , we always guess $3/8$. But $X \leq Y$, so a small observed y tells us that X must also be small. The estimate $\hat{x}_M(y)$ from (b) adjusts to this information instead of using the same guess in every situation. This helps reduce the average squared error, although it need not give a closer guess on every individual trial.

Question 2

Given $X = x$, the channel model becomes $Y = x + W$. Independence means that conditioning on $X = x$ does not change the Gaussian distribution of W . Adding the constant x shifts the mean but does not change the variance:

$$\mu_{Y|X=x} = x + E[W | X = x] = x + 0 = x, \\ \text{Var}[Y | X = x] = \text{Var}[W | X = x] = \sigma_W^2, \quad \sigma_{Y|X=x} = \sqrt{\sigma_W^2} = \sigma_W.$$

Thus the likelihood is the noise PDF shifted by x , $f_{Y|X}(y | x) = f_W(y - x)$. The Gaussian prior and conditional likelihood are

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-x^2/(2\sigma_X^2)}, \quad f_{Y|X}(y | x) = \frac{1}{\sqrt{2\pi}\sigma_W} e^{-(y-x)^2/(2\sigma_W^2)}.$$

(a) ML estimate.

$$\hat{x}_{\text{ML}}(y) = \arg \max_x f_{Y|X}(y | x).$$

Maximising the likelihood is equivalent to minimising $(y - x)^2$. Its derivative is $2(x - y)$, so the unique minimum is at $x = y$:

$$\boxed{\hat{x}_{\text{ML}}(y) = y}.$$

(b) MAP estimate.

Bayes' density formula is $f_{X|Y}(x | y) = f_{Y|X}(y | x)f_X(x)/f_Y(y)$. The observation y is fixed. Since the positive denominator $f_Y(y)$ does not depend on x , it does not affect which x maximises the posterior:

$$\hat{x}_{\text{MAP}}(y) = \arg \max_x f_{Y|X}(y | x)f_X(x).$$

Substitute both Gaussian PDFs and multiply:

$$f_{Y|X}(y | x)f_X(x) = \frac{1}{2\pi\sigma_W\sigma_X} e^{-(y-x)^2/(2\sigma_W^2)} e^{-x^2/(2\sigma_X^2)} \\ = \frac{1}{2\pi\sigma_W\sigma_X} \exp \left[- \left\{ \frac{(y-x)^2}{2\sigma_W^2} + \frac{x^2}{2\sigma_X^2} \right\} \right].$$

The *exponentials* multiply, so their *exponents add*: $e^u e^v = e^{u+v}$. The positive prefactor is constant with respect to x . Define the expression inside the braces as

$$g(x) = \frac{(y-x)^2}{2\sigma_W^2} + \frac{x^2}{2\sigma_X^2}.$$

The function e^{-g} decreases as g increases. Therefore the largest posterior occurs at the *smallest* $g(x)$: this is why maximisation becomes minimisation,

$$\hat{x}_{\text{MAP}}(y) = \arg \max_x e^{-g(x)} = \arg \min_x g(x).$$

Differentiate directly using the chain rule and set the derivative to zero:

$$g'(x) = -\frac{y-x}{\sigma_W^2} + \frac{x}{\sigma_X^2} \\ = \frac{x-y}{\sigma_W^2} + \frac{x}{\sigma_X^2} = 0, \\ \sigma_X^2(x-y) + \sigma_W^2 x = 0 \implies (\sigma_X^2 + \sigma_W^2)x = \sigma_X^2 y.$$

$$\boxed{\hat{x}_{\text{MAP}}(y) = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2} y}.$$

A separate second-derivative calculation is *not necessary here*: g is a sum of squared terms with positive weights, so it is strictly convex and this stationary point is its unique global minimum. As an optional check, $g''(x) = 1/\sigma_W^2 + 1/\sigma_X^2 > 0$ confirms this.

Question 2 (continued)

(c) Moments and the Gaussian MMSE estimate.

Independence of X and W gives $\text{Cov}[X, W] = 0$. Using this fact and linearity, with $\text{Cov}[X, X] = \text{Var}[X]$,

$$\begin{aligned}\mu_Y &= \text{E}[X + W] = \mu_X + \mu_W = 0, \\ \text{Var}[Y] &= \text{Var}[X] + \text{Var}[W] + 2\text{Cov}[X, W] = \sigma_X^2 + \sigma_W^2, \\ \text{Cov}[X, Y] &= \text{Cov}[X, X + W] \\ &= \text{Cov}[X, X] + \text{Cov}[X, W] \\ &= \text{Var}[X] + 0 = \sigma_X^2.\end{aligned}$$

The pair (X, Y) is jointly Gaussian because $Y = X + W$ and X, W are independent Gaussian variables. The MMSE estimate is the conditional mean. Apply the lecture's Gaussian conditional-mean formula, with $\rho_{X,Y} = \text{Cov}[X, Y]/(\sigma_X\sigma_Y)$:

$$\hat{x}_M(y) = \text{E}[X | Y = y] = \rho_{X,Y} \frac{\sigma_X}{\sigma_Y} (y - \mu_Y) + \mu_X = \frac{\text{Cov}[X, Y]}{\text{Var}[Y]} y = \boxed{\frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2} y}.$$

This is also the LMSE estimate, and its expression agrees with the MAP estimate in (b). No derivation of the conditional PDF or error variance is needed.

Question 3

(a) Unbiasedness.

Since $\text{E}[X_i] = 0$,

$$\text{E}[X_i^2] = \text{Var}[X_i] + (\text{E}[X_i])^2 = \sigma^2.$$

By linearity of expectation,

$$\text{E}[\hat{\sigma}_n^2] = \frac{1}{n} \sum_{i=1}^n \text{E}[X_i^2] = \frac{1}{n} (n\sigma^2) = \boxed{\sigma^2}.$$

The bias $\text{E}[\hat{\sigma}_n^2] - \sigma^2$ is therefore zero.

(b) Sample size for the required precision.

Standard error means the standard deviation of the *estimator*, not that of an individual scale error. Lecture 25 writes the standard error as $\sqrt{e_n}$, where e_n is the mean square error of an unbiased estimator. Here, by (a),

$$e_n = \text{E}[(\hat{\sigma}_n^2 - \sigma^2)^2] = \text{Var}[\hat{\sigma}_n^2] = \frac{2\sigma^4}{n}.$$

Thus the lecture notation and the abbreviation SE give the same quantity:

$$\text{SE}(\hat{\sigma}_n^2) = \sqrt{e_n} = \sqrt{\text{Var}[\hat{\sigma}_n^2]} = \sqrt{\frac{2\sigma^4}{n}} < \sqrt{\frac{2(0.5)^2}{n}} = \sqrt{\frac{0.5}{n}}.$$

Because the variance bound is strict, it suffices to require

$$\sqrt{\frac{0.5}{n}} \leq 0.08 \iff n \geq \frac{0.5}{(0.08)^2} = 78.125.$$

The smallest integer is therefore $\boxed{n = 79}$. For $n \leq 78$, an allowed variance sufficiently close to 0.5 fails the required precision, so a smaller sample does not give a guarantee. The standard-error unit is kg^2 , because the estimated quantity is a variance.

(c) Consistency.

Apply Chebyshev's inequality to the estimator, whose mean is σ^2 :

$$P(|\hat{\sigma}_n^2 - \sigma^2| \geq \varepsilon) \leq \frac{\text{Var}[\hat{\sigma}_n^2]}{\varepsilon^2} = \frac{2\sigma^4}{n\varepsilon^2} \longrightarrow 0 \quad (n \rightarrow \infty).$$

This holds for every fixed $\varepsilon > 0$, so $\hat{\sigma}_n^2$ converges in probability to σ^2 and is a consistent estimator.

Question 4

$H_0 = \{X = 0\}$ and $H_1 = \{X = 4\}$ describe the true voltage; the receiver observes $Y \in \{L, M, H\}$. Either hypothesis can produce any reading. The category H means “high”, not H_1 .

(a) Marginal PMF and posterior probability of H_1 .

Total probability and Bayes’ theorem give

$$P_Y(y) = \sum_{i=0}^1 P(Y = y | H_i)P(H_i), \quad P(H_1 | Y = y) = \frac{P(Y = y | H_1)P(H_1)}{P_Y(y)}.$$

y	$P_Y(y)$	$P(H_1 Y = y)$
L	$0.6(0.8) + 0.1(0.2) = 0.50$	$0.1(0.2) \div 0.50 = 0.04$
M	$0.3(0.8) + 0.4(0.2) = 0.32$	$0.4(0.2) \div 0.32 = 0.25$
H	$0.1(0.8) + 0.5(0.2) = 0.18$	$0.5(0.2) \div 0.18 \approx 0.5556$

(b) ML and MAP decisions.

Likelihoods are conditional probabilities: $P_{Y|H_i}(y) = P(Y = y | H_i)$. Using the lecture’s H_0/H_1 likelihood ratio:

ML : choose H_0 if $\frac{P_{Y|H_0}(y)}{P_{Y|H_1}(y)} \geq 1.00$; otherwise choose H_1 ,

MAP : choose H_0 if $\frac{P_{Y|H_0}(y)}{P_{Y|H_1}(y)} \geq \frac{P(H_1)}{P(H_0)} = 0.2 \div 0.8 = 0.25$; otherwise choose H_1 .

ML compares likelihoods; MAP uses the prior ratio to adjust the decision threshold.

y	Conditional likelihoods		Ratios		Decision		
	$P_{Y H_0}(y)$	$P_{Y H_1}(y)$	$\frac{P_{Y H_0}(y)}{P_{Y H_1}(y)}$	$\frac{P(H_1)}{P(H_0)}$	ML	MAP	MAP region
L	0.6	0.1	$0.6 \div 0.1 = 6.00$	0.25	H_0	H_0	A_0
M	0.3	0.4	$0.3 \div 0.4 = 0.75$	0.25	H_1	H_0	A_0
H	0.1	0.5	$0.1 \div 0.5 = 0.20$	0.25	H_1	H_1	A_1

For $Y = M$, $0.75 < 1.00$ but $0.75 > 0.25$: ML chooses H_1 , MAP chooses H_0 .

For the chosen rule, A_i contains readings assigned to decision H_i :

$$A_i = \{y \in S_Y : \text{the receiver chooses } H_i\}, \quad i = 0, 1.$$

Collecting the MAP rows gives

$$\boxed{\text{MAP: } A_0 = \{L, M\}, \quad A_1 = \{H\}.}$$

In probability expressions, A_i denotes the event of deciding H_i , not a claim that H_i is true.

Question 4 (continued)

(c) MAP error probabilities.

Using the MAP regions: a false alarm decides $X = 4$ when $X = 0$ is true; a miss decides $X = 0$ when $X = 4$ is true. The error cells from (b) are

y	MAP chooses	Error when	Conditional contribution
L	H_0	H_1 true: miss	$P(Y = L H_1) = 0.1$
M	H_0	H_1 true: miss	$P(Y = M H_1) = 0.4$
H	H_1	H_0 true: false alarm	$P(Y = H H_0) = 0.1$

Condition on the true hypothesis:

$$\begin{aligned}
 P_{\text{FA}} &= P(A_1 | H_0) = P(Y = H | H_0) = \boxed{0.1}, \\
 P_{\text{MISS}} &= P(A_0 | H_1) = P(Y = L | H_1) + P(Y = M | H_1) \\
 &= 0.1 + 0.4 = \boxed{0.5}, \\
 P_{\text{ERR}} &= P(H_0)P_{\text{FA}} + P(H_1)P_{\text{MISS}} = 0.8(0.1) + 0.2(0.5) = \boxed{0.18}.
 \end{aligned}$$

For interest, this is equivalently the total joint probability of the two error events:

$$P_{\text{ERR}} = P(A_1, H_0) + P(A_0, H_1).$$

(d) Unequal costs.

$C_{10} = 1$ is the false-alarm cost; $C_{01} = 4$ is the miss cost. Correct decisions cost zero. The minimum-cost rule is

$$\text{choose } H_0 \text{ if } \frac{P_{Y|H_0}(y)}{P_{Y|H_1}(y)} \geq \frac{P(H_1)C_{01}}{P(H_0)C_{10}} = [0.2(4)] \div [0.8(1)] = 1.00; \quad \text{otherwise choose } H_1.$$

Compare the ratios from (b) with 1.00, rather than MAP's 0.25:

y	Likelihood ratio	Cost-adjusted threshold	Minimum-cost decision
L	$6.00 \geq 1.00$	1.00	H_0
M	$0.75 < 1.00$	1.00	H_1
H	$0.20 < 1.00$	1.00	H_1

Now M leads to H_1 , so $A_0 = \{L\}$ and $A_1 = \{M, H\}$, giving $P_{\text{FA}} = 0.3 + 0.1 = 0.4$ and $P_{\text{MISS}} = 0.1$. The expected cost is

$$E[C] = P(A_1 | H_0)P(H_0)C_{10} + P(A_0 | H_1)P(H_1)C_{01}.$$

Substituting each rule's conditional error probabilities:

$$\begin{aligned}
 E[C]_{\text{min}} &= 0.4(0.8)(1) + 0.1(0.2)(4) = \boxed{0.40}, \\
 E[C]_{\text{MAP}} &= 0.1(0.8)(1) + 0.5(0.2)(4) = \boxed{0.48}.
 \end{aligned}$$

MAP minimises error probability; unequal costs can favour a different rule.

(e) Blind and observation-based MMSE estimates.

$$\hat{x}_B = E[X] = 0(0.8) + 4(0.2) = \boxed{0.8 \text{ V}}, \quad \hat{x}_M(y) = E[X | Y = y] = 0P(H_0 | y) + 4P(H_1 | y).$$

Using the posterior probabilities from (a), each estimate is:

$$\begin{aligned}
 \hat{x}_M(L) &= 0(0.96) + 4(0.04) = \boxed{0.16 \text{ V}}, \\
 \hat{x}_M(M) &= 0(0.75) + 4(0.25) = \boxed{1 \text{ V}}, \\
 \hat{x}_M(H) &= 0(0.08 \div 0.18) + 4(0.10 \div 0.18) = 0.40 \div 0.18 = \boxed{\frac{20}{9} \text{ V}} \approx 2.22 \text{ V}.
 \end{aligned}$$

A hypothesis test must choose a source symbol; an MMSE estimate instead minimises squared error and may lie between the two possible symbols.

Question 5

(a) Observed sample mean and variances.

Evaluating the estimators at the five given readings:

$$M_5 = \frac{1}{5} \sum_{i=1}^5 T_i = \frac{21 + 22 + 22 + 23 + 27}{5} = \boxed{23^\circ\text{C}}.$$

The squared deviations are

T_i	21	22	22	23	27	Total
$T_i - M_5$	-2	-1	-1	0	4	0
$(T_i - M_5)^2$	4	1	1	0	16	22

Therefore

$$V_5 = \frac{1}{5} \sum_{i=1}^5 (T_i - M_5)^2 = \frac{22}{5} = \boxed{4.4(^{\circ}\text{C})^2}, \quad V'_5 = \frac{1}{5-1} \sum_{i=1}^5 (T_i - M_5)^2 = \frac{22}{4} = \boxed{5.5(^{\circ}\text{C})^2}.$$

(b) Estimated standard error versus individual-reading standard deviation.

For independent readings, each with variance σ^2 ,

$$\begin{aligned} \text{Var}[M_n] &= \text{Var}\left[\frac{1}{n} \sum_{i=1}^n T_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[T_i] \\ &= \frac{1}{n^2} (n\sigma^2) = \frac{\sigma^2}{n}. \end{aligned}$$

Since M_n is unbiased, $e_n = \text{Var}[M_n]$. As in Question 3, standard error is $\sqrt{e_n}$:

$$\text{SE}(M_n) = \sqrt{e_n} = \sqrt{\text{Var}[M_n]} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}.$$

Estimate σ^2 using the unbiased $V'_5 = 5.5$ from (a), and set $n = 5$:

$$\widehat{\text{SE}}(M_5) = \sqrt{\frac{V'_5}{5}} = \sqrt{\frac{5.5}{5}} = \sqrt{1.1} \approx \boxed{1.049^\circ\text{C}}.$$

For an individual reading, the estimated standard deviation is

$$\hat{\sigma} = \sqrt{V'_5} = \sqrt{5.5} \approx \boxed{2.345^\circ\text{C}}.$$

Standard deviation describes the variability of individual readings; standard error describes the variability of their mean across repeated samples. Averaging five independent readings reduces the standard deviation by a factor of $\sqrt{5}$.

Question 6 (bonus)

(a) Chernoff bound.

For the supplied MGF, which is finite for all real s ,

$$\begin{aligned} P(X \geq 30) &\leq \min_{s \geq 0} e^{-30s} \phi_X(s) \\ &= \min_{s \geq 0} \exp\{-30s + 20(e^s - 1)\}. \end{aligned}$$

Because the exponential is increasing, minimise its exponent $h(s) = -30s + 20(e^s - 1)$:

$$h'(s) = -30 + 20e^s = 0 \implies e^{s^*} = \frac{3}{2} \implies s^* = \ln(3/2) > 0.$$

Also $h''(s) = 20e^s > 0$, so this is the global minimum. Substituting into the bound,

$$P(X \geq 30) \leq \exp\{-30 \ln(3/2) + 20(3/2 - 1)\} = \boxed{e^{10 - 30 \ln(3/2)} \approx 0.1149}.$$

(b) Markov comparison.

Since $X \geq 0$, Markov's inequality gives

$$P(X \geq c) \leq \frac{\text{E}[X]}{c} \implies P(X \geq 30) \leq \frac{20}{30} = \boxed{\frac{2}{3} \approx 0.6667}.$$

Chernoff is tighter here. Markov uses only the mean, whereas Chernoff uses the MGF and hence more distributional information. Both calculations give upper bounds: the inequalities do not assert equality with the actual event probability.

Question 7 (bonus)

(a) Blind estimation.

For a uniform RV on $[a, b]$, $E[T] = (a + b)/2$ and $\text{Var}[T] = (b - a)^2/12$. Therefore

$$\hat{t}_B = E[T] = \frac{20 + 30}{2} = \boxed{25^\circ\text{C}}, \quad e_B = E[(T - \hat{t}_B)^2] = \text{Var}[T] = \frac{(30 - 20)^2}{12} = \boxed{\frac{25}{3} (\text{°C})^2}.$$

(b) Alarm probabilities and conditional PDFs.

The original density is $f_T(t) = 1/10$ for $20 \leq t \leq 30$, and zero otherwise.

$$P(B) = P(T > 26) = \int_{26}^{30} f_T(t) dt = \frac{30 - 26}{10} = \boxed{0.4}, \quad P(\bar{B}) = 1 - P(B) = \boxed{0.6}.$$

Because the event depends only on T , restrict the marginal PDF to the appropriate temperature interval and divide by the event probability:

$$f_{T|B}(t) = \begin{cases} \frac{f_T(t)}{P(B)}, & 26 < t \leq 30, \\ 0, & \text{otherwise,} \end{cases} = \begin{cases} \frac{1}{4}, & 26 < t \leq 30, \\ 0, & \text{otherwise,} \end{cases}$$

$$f_{T|\bar{B}}(t) = \begin{cases} \frac{f_T(t)}{P(\bar{B})}, & 20 \leq t \leq 26, \\ 0, & \text{otherwise,} \end{cases} = \begin{cases} \frac{1}{6}, & 20 \leq t \leq 26, \\ 0, & \text{otherwise.} \end{cases}$$

Each conditional density integrates to one.

(c) Event-based MMSE estimates.

The conditional MMSE estimate is the conditional expected value:

$$\hat{t}_{\text{on}} = E[T | B] = \int_{-\infty}^{\infty} t f_{T|B}(t) dt = \frac{1}{4} \left[\frac{t^2}{2} \right]_{26}^{30} = \boxed{28^\circ\text{C}},$$

$$\hat{t}_{\text{off}} = E[T | \bar{B}] = \int_{-\infty}^{\infty} t f_{T|\bar{B}}(t) dt = \frac{1}{6} \left[\frac{t^2}{2} \right]_{20}^{26} = \boxed{23^\circ\text{C}}.$$

Equivalently, these are the midpoints of the two conditional uniform intervals.

(d) Conditional and average MSE.

At the conditional mean, the conditional MSE equals the conditional variance. Apply the uniform-variance formula to the intervals of lengths 4 and 6:

$$e_{\text{on}} = E[(T - \hat{t}_{\text{on}})^2 | B] = \text{Var}[T | B] = \frac{(30 - 26)^2}{12} = \boxed{\frac{4}{3}},$$

$$e_{\text{off}} = E[(T - \hat{t}_{\text{off}})^2 | \bar{B}] = \text{Var}[T | \bar{B}] = \frac{(26 - 20)^2}{12} = \boxed{3}.$$

The observer learns *either* alarm state and uses the corresponding estimate. The average error is therefore found by weighting the two conditional errors:

$$e_{\text{alarm}} = P(B)e_{\text{on}} + P(\bar{B})e_{\text{off}} = 0.4 \left(\frac{4}{3} \right) + 0.6(3) = \boxed{\frac{7}{3} (\text{°C})^2}.$$

Thus observing only the alarm state still reduces the average MSE from $25/3$ to $7/3$. The conditional errors above also have units $(\text{°C})^2$.