PROBLEM 1. \( p_{VW}(v, w) \).

(a)

\[
E[V + W] = \int \int (v + w)p_{VW}(v, w) \, dv \, dw \tag{1}
\]

\[
= \int \int (vp_{VW}(v, w) + wp_{VW}(v, w)) \, dv \, dw \tag{2}
\]

\[
= \int \int vp_{VW}(v, w) \, dv \, dw + \int \int wp_{VW}(v, w) \, dv \, dw \tag{3}
\]

\[
= \int v \int p_{VW}(v, w) \, dw \, dv + \int w \int p_{VW}(v, w) \, dv \, dw \tag{4}
\]

\[
= \int vp_{V}(v) \, dv + \int wp_{W}(w) \, dw \tag{5}
\]

\[
= E[V] + E[W] \tag{6}
\]

(b)

\[
E[V \cdot W] = \int \int (v \cdot w)p_{VW}(v, w) \, dv \, dw \tag{7}
\]

\[
= \int \int (v \cdot w)p_{V}(v)p_{W}(w) \, dv \, dw \tag{8}
\]

\[
= \int vp_{V}(v) \, dv \cdot \int wp_{W}(w) \, dw \tag{9}
\]

\[
= E[V] \cdot E[W] \tag{10}
\]

(c) Assume \( V = W \) and \( \Pr(V = 1) = \Pr(V = -1) = \frac{1}{2} \). We compute \( E[V] = E[W] = 0 \) and \( E[WW] = 1 \), so \( E[WW] \neq E[V]E[W] \).

Now suppose \((V, W)\) takes values of \((1, 1), (1, -1), (-1, 1), (-1, -1), (0, 0)\) with equal probability \(\frac{1}{5}\). Because \( \Pr(W = 0|V = 1) = 0 \neq \frac{1}{5} = \Pr(W = 0) \), \( V \) and \( W \) are not independent. We compute \( E[V] = E[W] = 0 \) and \( E[VW] = 0 \), so \( E[VW] = E[V]E[W] \).

(d) Assume that \( V \) and \( W \) are independent and let \( \sigma_{V}^{2} \) and \( \sigma_{W}^{2} \) be the variances of \( V \) and \( W \), respectively. Show that the variance of \( V + W \) is given by \( \sigma_{V+W}^{2} = \sigma_{V}^{2} + \sigma_{W}^{2} \).

\[
\sigma_{V+W}^{2} = E[(V + W)^2] - E[V + W]^2 \tag{11}
\]

\[
= E[V^2] + E[W^2] + 2E[VW] - (E[V] + E[W])^2 \tag{12}
\]

\[
\]

\[
= E[V^2] - E[V]^2 + E[W^2] - E[W]^2 \tag{14}
\]

\[
= \sigma_{V}^{2} + \sigma_{W}^{2} \tag{15}
\]
Problem 2.

(a)

\[
\sum_{n>0} \Pr(N \geq n) = \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \Pr(N = m) = \sum_{m=1}^{\infty} \sum_{n=1}^{m} \Pr(N = m) = \sum_{m=1}^{\infty} m \Pr(N = m) = E[N]
\]

(b)

\[
\int_0^{\infty} \Pr(x \geq a) \, da = \int_0^{\infty} \int_a^{\infty} f_x(t) \, dt \, da = \int_0^{\infty} \int_0^{t} f_x(t) \, da \, dt = \int_0^{\infty} t \, f_x(t) \, dt = E[X]
\]

(c) The main point is to note that \( G(t) = P(X \geq t) \) is a non-increasing function of \( t \). So for any fixed value of \( a > 0 \), the rectangle between point \((0, 0)\) and \((a, G(a))\) lies below the function \( G(t) \). In conclusion, it follows from the discussion above that

\[
aG(a) \leq \int_0^{a} G(a) \, dt \leq \int_0^{a} G(t) \, dt \leq \int_0^{\infty} G(t) \, dt,
\]

which means

\[
a \Pr(X \geq a) \leq E[X]
\]

(d) Assume

\[X = (Y - E[Y])^2 \quad X \geq 0\]

Using part (c), we have

\[a \Pr(X \geq a) \leq E[X].\]

Therefore, one could conclude that

\[a \Pr((Y - E[Y])^2 \geq a) \leq E((Y - E[Y])^2).\]

Setting \( b = \sqrt{a} \), we have

\[
\Pr(|Y - E[Y]| \geq b) = \Pr((Y - E[Y])^2 \geq b^2) \leq \frac{E((Y - E[Y])^2)}{b^2} = \frac{\sigma_Y^2}{b^2}.
\]

Problem 3.
(a) \( \Pr(X_1 \leq X_2) = \frac{1}{2} \). We know because of independence we have, \( f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2) \), and we want to find the probability of \( x_1 \) being minimum of two. This event partitions the probability space into two equal sub-sets, the other one is \( x_2 \) being the minimum of the two. The only problem is the boundary line \( x_1 = x_2 \), which we assume is a part of first sub-set, but because \( f_x \) is a continuous random variable the line \( x_1 = x_2 \) has zero probability mass and because \( f_{X_1}(x_1)f_{X_2}(x_2) \) is symmetric with respect to the line \( x_1 = x_2 \), we conclude that the event \( \min(x_1, x_2) = x_1 \) partitions the whole probability space into two equally probable regions.

(b) \( \Pr(X_1 \leq X_2; X_1 \leq X_3) = \frac{1}{3} \): We follow the exact same argument as the part (a), this time the probability space is partitioned into three equally probable sub-sets, in each of sub-sets one of the three random variable is minimum.

(c) Similar to last parts, we can show that

\[
\Pr(X_1 \leq X_2; X_1 \leq X_3; \ldots; X_1 \leq X_{n-1}; X_1 \leq X_n) = \frac{1}{n}
\]

and

\[
\Pr(X_1 \leq X_2; X_1 \leq X_3; \ldots; X_1 \leq X_{n-1}) = \frac{1}{n-1}
\]

We know

\[
\Pr(N = n) = \Pr(X_1 \leq X_2; X_1 \leq X_3; \ldots; X_1 \leq X_{n-1}; X_1 > X_n) = \Pr(X_1 \leq X_2; X_1 \leq X_3; \ldots; X_1 \leq X_{n-1}) - \Pr(X_1 \leq X_2; X_1 \leq X_3; \ldots; X_1 \leq X_{n-1}; X_1 \leq X_n)
\]

\[
= \frac{1}{n-1} - \frac{1}{n} = \frac{1}{n^2 - n}, \quad n > 1
\]

Using properties of telescopic series, we conclude

\[
\Pr(N \geq n) = \sum_{m=n}^{\infty} \Pr(N = m) = \frac{1}{n-1} - \frac{1}{n} + \frac{1}{n-1} - \frac{1}{n+1} + \ldots
\]

\[
= \frac{1}{n-1}, \quad n \geq 2
\]

(d) We use part (a) of Problem 2.

\[
E(N) = \sum_{n=0}^{\infty} \Pr(N \geq n) = \sum_{n=1}^{\infty} \frac{1}{n-1} \to \infty
\]

(We know that series \( \frac{1}{n} \) is divergent.)

(e) The symmetry of the \( f_{X_1}(x_1)f_{X_2}(x_2) \) still holds because of independence but in the discrete case it is possible to put some probability mass on the line \( x_1 = x_2 \). Therefore in the discrete case the event \( x_1 \leq x_2 \) does not partition the whole probability space into two equally probable sub-spaces. The same as before we can conclude that \( \Pr(X_1 < X_2) = \Pr(X_2 < X_1) \). We know \( \Pr(X_1 < X_2) + \Pr(X_1 = X_2) + \Pr(X_2 < X_1) = 1 \). From these two we conclude that \( \Pr(X_1 \leq X_2) \geq \frac{1}{2} \). Similarly we conclude that

\[
\Pr(X_1 \leq X_2; X_1 \leq X_3; \ldots; X_1 \leq X_{n-1}; X_1 \leq X_n) \geq \frac{1}{n}.
\]
Following the steps in part (d), we can show that
\[ E(N) \geq \sum_{n>1} \frac{1}{n-1} \rightarrow \infty \]

**Problem 4.** Let’s consider the case where \( n = 2 \) first, we have
\[ P(Z = 0) = P(X_1 \oplus X_2 = 0) = P(X_1 = 0, X_2 = 0) + P(X_1 = 1, X_2 = 1) = \frac{1}{2}, \]
in which we used independence of \( X_1 \) and \( X_2 \).
By induction, one could easily show that for arbitrary \( n \), we have
\[ P(Z = 0) = \frac{1}{2} . \]

(a) \[ P(Z = z | X_1 = x_1) = P(X_1 \oplus X_2 \oplus \cdots \oplus X_n = z | X_1 = x_1) \]
\[ = P(X_2 \oplus \cdots \oplus X_n = z \oplus x_1 | X_1 = x_1) \]
\[ = P(X_2 \oplus \cdots \oplus X_n = z \oplus x_1) \]
\[ = \frac{1}{2} = P(Z = z) \]
in (32) we used that \( X_i \)’s are independent. We conclude that \( Z \) is independent of \( X_1 \)

(b) \[ P(Z = z | X_1, \ldots, X_{n-1} = x_1, \ldots, x_{n-1}) = \]
\[ P(X_1 \oplus X_2 \oplus \cdots \oplus X_n = z | X_1, \ldots, X_{n-1} = x_1, \ldots, x_{n-1}) = \]
\[ P(X_n = z \oplus x_1 \oplus \cdots \oplus x_{n-1} | X_1, \ldots, X_{n-1} = x_1, \ldots, x_{n-1}) = \]
\[ = \frac{1}{2} = P(Z = z) \]
in (37) we used that \( X_i \)’s are independent. We conclude that \( Z \) is independent of \( X_1, \ldots, X_{n-1} \).

(c) No, \( Z \) is a deterministic function of \( X_1, \ldots, X_n \), which means
\[ P(Z = z | X_1, \ldots, X_n = x_1, \ldots, x_n) \]
is either 0 or 1 depending on the values of \( x_1, \ldots, x_n \) and \( z \).

(d) Suppose \( \Pr(X_i = 1) = \frac{3}{4} \), we have
\[ P(Z = 0) = P(X_1 \oplus X_2 = 0) = P(X_1 = 0, X_2 = 0) + P(X_1 = 1, X_2 = 1) = \frac{9 + 1}{16} = \frac{5}{8} , \]
but
\[ P(Z = 0 | X_1 = 0) = P(X_1 \oplus X_2 = 0 | X_1 = 0) \]
\[ = P(X_2 = 0 | X_1 = 0) \]
\[ = \frac{1}{4} \neq \frac{5}{8} = P(Z = 0), \]
in which we used that \( X_1 \) and \( X_2 \) are independent. We conclude that \( Z \) is not independent of \( X_1 \).
Problem 5. (1) Let $D_0, D_1$ be the MAP decision regions for hypotheses 0 and 1 when the a-priori probabilities are $(\pi_0, 1 - \pi_0)$. Similarly, let $D_0', D_1'$ be the MAP decision regions for hypotheses 0 and 1 when the a-priori probabilities are $(\pi_0', 1 - \pi_0')$, and $D_0'', D_1''$ be the MAP decision regions for hypotheses 0 and 1 when the a-priori probabilities are $(\pi_0'', 1 - \pi_0'')$, where $\pi_0'' = \lambda \pi_0 + (1 - \lambda) \pi_0'$. Thus

\[
V(\pi_0) = \pi_0 p_0(D_1) + (1 - \pi_0) p_1(D_0),
\]
\[
V(\pi_0') = \pi_0 p_0(D_1') + (1 - \pi_0) p_1(D_0'),
\]
\[
V(\pi_0'') = \pi_0 p_0(D_1'') + (1 - \pi_0) p_1(D_0'').
\]

(2) Since the MAP rule minimizes the error probability, using any other decision regions in any of the above will increase the probability of error. So,

\[
V(\pi_0) \leq \pi_0 p_0(D_0'') + (1 - \pi_0) p_1(D_0''),
\]
\[
V(\pi_0') \leq \pi_0' p_0(D_0'') + (1 - \pi_0') p_1(D_0'').
\]

Multiplying the first by $\lambda$ and the second by $(1 - \lambda)$ and adding we get the desired result:

\[
\lambda V(\pi_0) + (1 - \lambda) V(\pi_0') \leq (\lambda \pi_0 + (1 - \lambda) \pi_0') p_0(D_1'') + (1 - (\lambda \pi_0 + (1 - \lambda) \pi_0')) p_1(D_0'')
\]
\[
= V(\lambda \pi_0 + (1 - \lambda) \pi_0').
\]

Problem 6. We define

\[C(x_i) = 2\sigma^2 \log \Pr(x_i)\]

It is easy to show that for the optimal decision maker (MAP) in Gaussian noise, the detector finds $x_i$ so that

\[\langle x_i, x_i \rangle - 2\langle y, x_i \rangle - C(x_i)\]

is minimized.

We know the following for any $j \neq i$

\[
\langle x_i, x_i \rangle - 2\langle y_1, x_i \rangle - C(x_i) \leq \langle x_j, x_j \rangle - 2\langle y_1, x_j \rangle - C(x_j) \tag{44}
\]
\[
\langle x_i, x_i \rangle - 2\langle y_2, x_i \rangle - C(x_i) \leq \langle x_j, x_j \rangle - 2\langle y_2, x_j \rangle - C(x_j). \tag{45}
\]

Now let us consider the following,

\[
\langle x_i, x_i \rangle - 2\langle \alpha y_1 + (1-\alpha)y_2, x_i \rangle - C(x_i) = \langle x_i, x_i \rangle - 2\alpha \langle y_1, x_i \rangle - 2(1-\alpha) \langle y_2, x_i \rangle - C(x_i)
\]
\[
= \alpha \left[ \langle x_i, x_i \rangle - 2\langle y_1, x_i \rangle - C(x_i) \right] + (1-\alpha) \left[ \langle x_i, x_i \rangle - 2\langle y_2, x_i \rangle - C(x_i) \right]
\]
\[
\leq \alpha \left[ \langle x_j, x_j \rangle - 2\langle y_1, x_j \rangle - C(x_j) \right] + (1-\alpha) \left[ \langle x_j, x_j \rangle - 2\langle y_2, x_j \rangle - C(x_j) \right].
\]

In the last step we used 44 and 45. We conclude

\[
\langle x_i, x_i \rangle - 2\langle \alpha y_1 + (1-\alpha)y_2, x_i \rangle - C(x_i) \leq \langle x_j, x_j \rangle - 2\langle \alpha y_1 + (1-\alpha)y_2, x_j \rangle - C(x_j)
\]

for all $j \neq i$. Therefore, the decoder decodes $\alpha y_1 + (1-\alpha)y_2$ as $x_i$. 

5