1. Suppose that $X$ and $Y$ are independent random variables with finite variances such that $E(X) = E(Y)$. Show that

$$E[(X - Y)^2] = Var(X) + Var(Y)$$

Hint: $Var(X) = E[X^2] - (E[X])^2$ \[4 \text{ points}\]


$E[X^2] - 2(E[X])^2 + E[X^2]$ \[because $E(X)=E(Y)$\]

$E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2$ \[because $E(X)=E(Y)$\]

$Var(X) + Var(Y)$ \[because $Var(X)=E[X^2]-(E[X])^2$\]

2. A fair coin is tossed until a head is obtained. What is the expected number of tosses that will be required? \[5 \text{ points}\]

Let $X_t = 0$ be the event that you get a tails on round $t$

Let $X_t = 1$ be the event that you get a heads on round $t$

Let $Y$ be the random variable corresponding to the number of tosses before you get a heads

You need to compute:

$$\sum_{t=1}^{\infty} Y \times P(X_1 = 0, X_2 = 0, ..., X_t = 1)$$

This looks like:

$$P(X_1 = 1) \times 1 +$$

$$P(X_1 = 0, X_2 = 1) \times 2 +$$

$$P(X_1 = 0, X_2 = 0, X_3 = 1) \times 3 + ...$$

$$= \sum_{t=1}^{\infty} t \times (0.5)^t = \sum_{t=1}^{\infty} t \times (0.5)^{t-1} (0.5) = (0.5) \sum_{t=1}^{\infty} t \times (0.5)^{t-1}$$

$$= \frac{0.5}{(1-0.5)^2} = 2$$

3. Let $X$ be drawn from a Normal distribution $N(0,1)$. Let $Y = e^X$. What is the expectation $E[Y]$ and Variance $Var(Y)$? \[5 \text{ points}\]

(Hint: Use the completing the square trick)
4a) We would like to evaluate if coins that come from a nearby mint are fair or unfair (we will call this quality their fairness). Fair coins have an equal probability of landing heads or tails when flipped while unfair coins come up heads with probability 0.75. For simplicity, assume the coin does not land on its edge. Before flipping a coin, your prior belief is that the mint is fair with probability 0.5. You take three coins and flip them all simultaneously. They all come up heads. What is the probability that the mint produces fair coins given the evidence you just observed from the three coins? You may assume that the coin flips are all conditionally independent given their fairness. [5 points]

Putting this together,

\[ P(Fair = True \mid F_1 = H, F_2 = H, F_3 = H) = \frac{P(F_1 = H, F_2 = H, F_3 = H \mid Fair = True)P(Fair = True)}{P(F_1 = H, F_2 = H, F_3 = H)} \]

\[ = \frac{P(F_1 = H \mid Fair = True)P(F_2 = H \mid Fair = True)P(F_3 = H \mid Fair = True)}{P(F_1 = H, F_2 = H, F_3 = H)} \]

\[ = \frac{(0.5)(0.5)(0.5)}{0.2736} \]

\[ = 0.2284 \]
\[ P(Fair = True \mid F1 = H, F2 = H, F3 = H) \]
\[ = \frac{P(F1 = H, F2 = H, F3 = H \mid Fair = True)P(Fair = True)}{P(F1 = H, F2 = H, F3 = H)} \]
\[ = \frac{P(F1 = H \mid Fair = True)P(F2 = H \mid Fair = True)P(F3 = H \mid Fair = True)P(Fair = True)}{P(F1 = H, F2 = H, F3 = H)} \]

Calculating \( P(F1 = H, F2 = H, F3 = H) \)
\[ = P(F1 = H, F2 = H, F3 = H, Fair = True) + P(F1 = H, F2 = H, F3 = H, Fair = False) \]
\[ = P(F1 = H \mid Fair = True)P(F2 = H \mid Fair = True)P(F3 = H \mid Fair = True)P(Fair = True) + \]
\[ P(F1 = H \mid Fair = False)P(F2 = H \mid Fair = False)P(F3 = H \mid Fair = False)P(Fair = False) \]
\[ = (0.5)(0.5)(0.5)(0.5) + (0.75)(0.75)(0.75)(0.5) \]
\[ = 0.0625 + 0.211 \]
\[ = 0.2736 \]

Putting this together,
\[ P(Fair = True \mid F1 = H, F2 = H, F3 = H) \]
\[ = \frac{P(F1 = H, F2 = H, F3 = H \mid Fair = True)P(Fair = True)}{P(F1 = H, F2 = H, F3 = H)} \]
\[ = \frac{P(F1 = H \mid Fair = True)P(F2 = H \mid Fair = True)P(F3 = H \mid Fair = True)P(Fair = True)}{P(F1 = H, F2 = H, F3 = H)} \]
\[ = \frac{(0.5)(0.5)(0.5)(0.5)}{0.2736} \]
\[ = 0.2284 \]

b) Now suppose part (a) never happened and you initially believe the coins have equal probability of being fair or unfair. You flip the first coin and observe that it lands heads side up. What is your belief regarding the fairness of the coins given the evidence from this coin flip? [2 points]
\( P(Fair = True \mid F1 = H) \)
\[ = \frac{P(F1 = H \mid Fair = True)P(Fair = True)}{P(F1 = H \mid Fair = True)P(Fair = True) + P(F1 = H \mid Fair = False)P(Fair = False)} \]
\[ = \frac{(0.5)(0.5)}{(0.5)(0.5) + (0.75)(0.5)} = \frac{0.25}{0.25 + 0.375} = \frac{0.25}{0.625} = 0.4 \]

c) Having observed the first coin flip, you now flip the second coin and observe that it lands heads side up. Using the posterior probability from part (b) as your prior ie. \( P(Fair = True) = your\ answer\ from\ part\ (b)\), calculate your updated belief regarding the fairness of the coins given the second coin flip (Note: if you use the posterior from part (b) as
your prior, you only need to condition on the second coin flip, not both the first and second coin flips). [2 points]

\[
P(Fair = True \mid F2 = H) = \frac{P(F2 = H \mid Fair = True)P(Fair = True)}{P(F2 = H \mid Fair = True)P(Fair = True) + P(F2 = H \mid Fair = False)P(Fair = False)} = \frac{(0.5)(0.4)}{(0.5)(0.4) + (0.75)(0.6)} = \frac{0.2}{0.2 + 4.5} = \frac{0.2}{0.65} = 0.31
\]

d) Having observed the second coin flip, you now flip the third coin and observe that it lands heads side up. Using the posterior probability from part (c) as your prior, calculate your updated belief regarding the fairness of the coins based on the third coin flip. As in part (c), you only need to condition on the third coin flip, not all three. Is this final answer the same as the answer in part (a) when you flipped the coins simultaneously? [2 points]

\[
P(Fair = True \mid F3 = H) = \frac{P(F3 = H \mid Fair = True)P(Fair = True)}{P(F3 = H \mid Fair = True)P(Fair = True) + P(F3 = H \mid Fair = False)P(Fair = False)} = \frac{(0.5)(0.31)}{(0.31)(0.31) + (0.75)(0.69)} = \frac{0.155}{0.155 + 0.5175} = \frac{0.155}{0.6725} = 0.230
\]

5a. Show that, if all the variables \( C, X_1, \ldots, X_m \) are all binary-valued, then a naïve Bayes classifier can be written as the following linear function:

\[
\log \frac{P(C = 1 \mid X_1, \ldots, X_m)}{P(C = 0 \mid X_1, \ldots, X_m)} = \sum_{i=1}^{m} \beta_i X_i + \beta_0
\]

This linear function is exactly the same form as that used by logistic regression. [5 points]

Hint: You’ll probably be able to convert the log probability ratio into a linear combination of terms but will get stuck with trying to figure out how the \( X_i \)'s fit in. Note that \( X_i = 0 \) or \( 1 \). In order to get the linear function in the form above, you’ll need to use a simple math trick that cancels out a term (call it \( \alpha_{i,0} \)) when \( X_i = 1 \) and cancels out a term (call it \( \alpha_{i,1} \)) when \( X_i = 0 \).

\[
\log \frac{P(C = 1 \mid X_1, \ldots, X_m)}{P(C = 0 \mid X_1, \ldots, X_m)} = \log \frac{P(X_1, \ldots, X_m \mid C = 1)P(C = 1)}{P(X_1, \ldots, X_m \mid C = 0)P(C = 0)} = \log \frac{P(X_1, \ldots, X_m \mid C = 1)P(C = 1)}{P(X_1, \ldots, X_m \mid C = 0)P(C = 0)}
\]

\[
= \log \frac{P(C = 1) \prod_{i=1}^{m} P(X_i \mid C = 1)}{P(C = 0) \prod_{i=1}^{m} P(X_i \mid C = 0)} = \log \frac{P(C = 1)}{P(C = 0)} + \sum_{i=1}^{m} \log \frac{P(X_i \mid C = 1)}{P(X_i \mid C = 0)}
\]
Define

\[ \alpha_{i,0} = \log \frac{P(X_i = 0 \mid C = 1)}{P(X_i = 0 \mid C = 0)} \]
\[ \alpha_{i,1} = \log \frac{P(X_i = 1 \mid C = 1)}{P(X_i = 1 \mid C = 0)} \]

\[ = \sum_{j=1}^{m} \log(\alpha_{i,1} - \alpha_{i,0}) x_j + (\sum_{i=1}^{m} \alpha_{i,0} + \log \frac{P(C = 1)}{P(C = 0)}) \]