Part I

Below is a table specifying probabilities for a jointly discrete distribution for a random vector $(X,Y)'$. Use it throughout Part I of this question.

Table 1: Values of $f(x,y) = P[X = x, Y = y]$

y \ x	1	2	3	4
4	.01	.02	.04	.1
3	.02	.07	.2	.04
2	.04	.2	.07	.02
1	.1	.04	.02	.01

1. Find a function $h(x)$ such that $\mathrm{E}(Y - h(X))^2$ is minimum. (You need only find numerical values of $h(x)$ for $x = 1,2,3,4$.)

2. Find the distribution of $Z = X + Y$.

3. For $Z_1, Z_2, \ldots, Z_{25}$ iid with the distribution from problem **2**, approximate $P\left[\sum_{i=1}^{25} Z_i \le 130.0\right]$.

Suppose now that $(X_1,Y_1)', (X_2,Y_2)', \ldots, (X_{25},Y_{25})'$ are iid from the distribution in **Table 1**.

4. A bivariate central limit theorem implies that $(\bar{X},\bar{Y})'$ is approximately bivariate normal.

a) Identify constants $c_1, c_2,$ and c_3 so that

$$c_1\left(\bar{X} - \mathrm{E}\,X\right)^2 + c_2\left(\bar{X} - \mathrm{E}\,X\right)\left(\bar{Y} - \mathrm{E}\,Y\right) + c_3\left(\bar{Y} - \mathrm{E}\,Y\right)^2 \mathrel{\dot\sim} \chi^2_2 \quad .$$

b) Identify a mean and standard deviation for an approximately normal distribution for

$$R = \bar{X} / \bar{Y} .$$

Part II

Consider two jointly continuous distributions for variables $U,V,$ and W with densities on $[0,1]^3$ respectively

$$f_0(u,v,w)=1 \quad \text{and} \quad f_1(u,v,w)=\frac{2}{3}(u+v+w) \quad .$$

5. With parameter $\theta \in \Theta = \{0,1\}$ the small parametric family of distributions on $[0,1]^3$ with densities f_θ has a 1-dimensional sufficient statistic. Identify such a statistic and argue carefully that it is indeed sufficient.

We will here consider a "classification" problem that amounts to deciding between the possibilities $\theta = 0$ and $\theta = 1$ based on an observation (U,V,W). (In other terms, we consider a version of a simple versus simple hypothesis testing problem based on (U,V,W).) Suppose that with probability $\gamma \in (0,1)$ the parameter θ takes the value 1. (In Bayesian language, γ is a prior probability the observation comes from density f_1.)

6. Carefully identify a minimum error rate classifier, i.e. a function $\delta_\gamma : [0,1]^3 \to \{0,1\}$ with minimum value of

$$(1-\gamma) P_0\left[\delta(U,V,W)=1\right] + \gamma P_1\left[\delta(U,V,W)=0\right] \quad .$$

(Hint: This is a test/classifier δ minimizing the Bayes risk under the loss

$$L(\theta,\delta(u,v,w)) = I[\theta=0]\, I\left[\delta(u,v,w)=1\right] + I[\theta=1]\, I\left[\delta(u,v,w)=0\right] \quad .$$

It is also a Neyman-Pearson test for a particular choice of "k".)

7. If one can only observe the first co-ordinate of the vector (U,V,W) one could nevertheless attempt to do minimum error rate classification among those classifiers that are functions of U alone. Identify a function $\delta_\gamma^* : [0,1] \to \{0,1\}$ with minimum value of

$$(1-\gamma) P_0\left[\delta(U)=1\right] + \gamma P_1\left[\delta(U)=0\right] \quad .$$

For the case of $\gamma = .5$, it turns out that $\delta^*_{.5}(u) = I[u > .5]$ (that is, one decides in favor of $\theta = 1$ if and only if what's observed is larger than $.5$). In addition to classifier $\delta^*_{.5}(U)$, consider the classifiers $\delta^*_{.5}(V)$ and $\delta^*_{.5}(W)$. (One might suppose that they are somehow available for use from other sources than a source that has access to U.) **Table 2** gives (joint) probabilities for possible values of $\delta^*_{.5}(U), \delta^*_{.5}(V)$, and $\delta^*_{.5}(W)$ under the $\theta = 0$ and $\theta = 1$ models.

Table 2: (Joint) Probabilities for Vectors of Three (Marginal) Classifiers Under the $\theta = 0$ and $\theta = 1$ **Models**

$\theta = 0$

$\delta^*_{.5}(U)$	$\delta^*_{.5}(V)$	$\delta^*_{.5}(W)$	Prob
0	0	0	$1/8$
1	0	0	$1/8$
0	1	0	$1/8$
1	1	0	$1/8$
0	0	1	$1/8$
1	0	1	$1/8$
0	1	1	$1/8$
1	1	1	$1/8$

$\theta = 1$

$\delta^*_{.5}(U)$	$\delta^*_{.5}(V)$	$\delta^*_{.5}(W)$	Prob
0	0	0	$3/48$
1	0	0	$5/48$
0	1	0	$5/48$
1	1	0	$7/48$
0	0	1	$5/48$
1	0	1	$7/48$
0	1	1	$7/48$
1	1	1	$9/48$

The problem of "classifier fusion" is that of trying to base a classification decision not on the original data, but on only the output of some given set of classifiers.

8. If $\gamma = .5$ and one has access to the values of $\delta^*_{.5}(U), \delta^*_{.5}(V)$, and $\delta^*_{.5}(W)$, how should one make classification decisions? That is, how should one choose $T : \{0,1\}^3 \to \{0,1\}$ to minimize

$$(1-\gamma) P_0\left[T\left(\delta^*_{.5}(U), \delta^*_{.5}(V),\ \delta^*_{.5}(W)\right) = 1\right] + \gamma P_0\left[T\left(\delta^*_{.5}(U), \delta^*_{.5}(V),\ \delta^*_{.5}(W)\right) = 0\right] \ ?$$

That is, what is the minimum error rate function, T, of $\left(\delta^*_{.5}(U), \delta^*_{.5}(V), \delta^*_{.5}(W)\right)$? (For each 3-vector of 0's and 1's give a corresponding element of $\Theta = \{0,1\}$ for the value of T.) Explain.

9. Find the error rate for your classifier from problem **8**. Do you expect it to be *larger* or to be *smaller* than the error rate for $\delta_{.5}(U,V,W)$ from problem **6**? (Say which and why.)