Solution to Exercise 4.1.4: OK, the random observable quantity and the family of possible distributions are already given, call them \(X\) and \(\{P_\theta : \theta \in \Theta\}\), respectively. The kind of action we want to take is estimate a set, so let the action space consist of \(A = B_p\), the \(p\)-dimensional Borel sets, where \(p\) is the dimension of the parameter vector \(\theta\). We restrict attention to Borel sets since we know they are Lebesgue measurable, and our next step is to define the loss function \(L(\theta, A) = m(A)\), where \(m\) is Lebesgue measure on \(\mathbb{R}^p\). We require that the coverage probability of the confidence set is at least \(1 - \alpha\), so we define the space of allowable decision rules

\[
D = \{\delta : \Xi \rightarrow A : P_\theta[\theta \in \delta(X)] \geq 1 - \alpha\}, \quad \forall \theta \in \Theta.
\]

That seems to be about the sum of this problem.

Solution to Exercise 4.2.5: Exercise 2.3.10 was assigned in Stat 532 last semester, so we will just go through the solution there and pick out what we need for this exercise. There is very little additional work to do. We will use the notations from the solution that was given last semester.

For the Poisson family, the sufficient statistic is \(T(x) = x\), and the family is full rank, so this is minimal sufficient.

For the Binomial family, again, \(T(x) = x\), and the family is full rank, so again it is minimal sufficient.

For the Beta family, the sufficient statistic is \(T(x) = (\log x, \log(1 - x))\). Again, it is full rank, so minimal sufficient.

For the negative binomial, \(T(x) = x\) is again minimal sufficient for the same reasons as the previous examples.

Solution to Exercise 4.2.16: Letting \(X\) and \(Y\) denote the two independent random vectors corresponding to the two populations, it is clear that we have an exponential family here under all circumstances. So, we can apply the Proposition 4.2.5 to deliver the results on minimal sufficiency.

For the first model where all parameters are unrestricted,

\[
f_{\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2}(x, y) = \exp \left[ -\frac{1}{2\sigma_X^2} \sum_i x_i^2 + \frac{\mu_X}{\sigma_X^2} \sum_i x_i + \frac{1}{2\sigma_Y^2} \sum_j y_j^2 + \frac{\mu_Y}{\sigma_Y^2} \sum_j y_j - B(\theta) \right] h(x, y).
\]

Here, we don’t care about the \(B(\theta)\) and \(h(x, y)\). Our natural parameter and sufficient statistic vectors are given by

\[
T = \left( \sum_i x_i^2, \sum_i x_i, \sum_j y_j^2, \sum_j y_j \right)
\]

\[
\eta = \left( -\frac{1}{2\sigma_X^2}, \frac{\mu_X}{\sigma_X^2}, -\frac{1}{2\sigma_Y^2}, \frac{\mu_Y}{\sigma_Y^2} \right).
\]
Note that the space of values of the natural parameter vector is \((-\infty, 0) \times (-\infty, \infty) \times (-\infty, 0) \times (-\infty, \infty)\), which is a nonempty open set, so has nonempty interior. It was mentioned in lecture that the Cartesian product of sets with nonempty interior has nonempty interior. Let’s verify that. To say a set \(A\) has nonempty interior means it has a nonempty open subset, i.e. that there is \(x \in A\) such that there is an \(\epsilon > 0\) such that for all \(x_1\) such that \(\|x_1 - x\| < \epsilon\), we have \(x_1 \in A\). The set \(\mathcal{N}(x, \epsilon) = \{x_1 : \|x_1 - x\| < \epsilon\}\) is an \(\epsilon\)-neighborhood of \(x\), or just a neighborhood. So, \(A\) having nonempty interior means there is some element with a neighborhood contained in \(A\). Now if two sets \(A\) and \(B\) each have nonempty interior, then clearly their Cartesian product will have nonempty interior provided the product of two neighborhoods again contains a neighborhood. Hopefully it is clear that the Cartesian product of two subsets is again a subset (of the product set, of course). Also, if we have that \(\|(x_1, y_1) - (x, y)\| < \epsilon\), then clearly \(\|x_1 - x\| < \epsilon\) and \(\|y_1 - y\| < \epsilon\), so \(\mathcal{N}((x, y), \epsilon) \subset \mathcal{N}(x, \epsilon) \times \mathcal{N}(y, \epsilon)\).

Note that we need to verify the identifiability condition, which also implies that \(T\) does not satisfy any linear constraints. We know that two normal densities are not the same unless they have the same parameters, and they will be equal at most two points if they have different parameters (set them equal, take logarithms, simplify and the points where they are equal satisfy a polynomial root equation of degree at most 2). Hence, they give different measures, and this follows for products of marginal normal distributions.

In conclusion, for model (i), a minimal sufficient statistic is

\[
T = \left( \sum_i x_i^2, \sum_i x_i, \sum_j y_j^2, \sum_j y_j \right)
\]

Turning to model (ii), assume \(\sigma_X^2 = \sigma_Y^2 = \sigma^2\). Then the exponential family representation takes the form

\[
f_{\mu_X, \mu_Y, \sigma^2}(x, y) = \exp\left[\frac{-1}{2\sigma^2} \left( \sum_i x_i^2 + \sum_j y_j^2 \right) + \frac{\mu_X}{\sigma^2} \sum_i x_i + \frac{\mu_Y}{\sigma^2} \sum_j y_j - B(\theta) \right] h(x, y).
\]

The sufficient statistic and natural parameter vectors are given by

\[
T = \left( \sum_i x_i^2 + \sum_j y_j^2, \sum_i x_i, \sum_j y_j \right)
\]

\[
\eta = \left( \frac{-1}{2\sigma^2}, \frac{\mu_X}{\sigma^2}, \frac{\mu_Y}{\sigma^2} \right).
\]

The space for the natural parameter values is \((-\infty, 0) \times (-\infty, \infty) \times (-\infty, \infty)\), which has nonempty interior, so \(T\) is minimal sufficient here. One can verify that this model is identifiable similarly to the previous case. (This means \(T\) does not satisfy any linear constraints, so the family is full rank, and it follows \(T\) is complete, as well.)
For model (iii), the exponential family has the form

\[ f_{\mu, \sigma_X^2, \sigma_Y^2}(x, y) = \exp \left[ -\frac{1}{2\sigma_X^2} \sum_i x_i^2 + \frac{\mu}{\sigma_X^2} \sum_i x_i + -\frac{1}{2\sigma_Y^2} \sum_j y_j^2 + \frac{\mu}{\sigma_Y^2} \sum_j y_j - B(\theta) \right] h(x, y). \]

natural parameter and sufficient statistic vectors are given by

\[ T = \left( \sum x_i^2, \sum x_i, \sum y_j^2, \sum y_j \right) \]

\[ \eta = \left( -\frac{1}{2\sigma_X^2}, \frac{\mu}{\sigma_X^2}, -\frac{1}{2\sigma_Y^2}, \frac{\mu}{\sigma_Y^2} \right). \]

The space of values of \( \eta \) is contained in \(( -\infty, 0) \times ( -\infty, \infty) \times ( -\infty, 0) \times ( -\infty, \infty) \), but of course we cannot expect that the map \( \eta(\mu, \sigma_X^2, \sigma_Y^2) \) can fill out all of this 4-dimensional set. (It is mathematically possible to map a 3-dimensional set onto a 4-dimensional set with nonempty interior, but you can’t write the function as a simple formula.) In particular, we see that

\[ \eta_2 = -2\mu \eta_1, \quad \eta_4 = -2\mu \eta_3, \]

so points in \( \eta(\Theta) \) must satisfy the (nonlinear) constraint

\[ \eta_1 \eta_4 = \eta_2 \eta_3. \]

Note that both sides of this equation are 0 if and only if \( \mu = 0 \) (since \( \eta_1 \) and \( \eta_3 \) are nonzero, and if both sides are nonzero, then \( \mu = -\eta_2/(2\eta_1) = -\eta_4/(2\eta_3) \)). Thus, any points \((\eta_1, \eta_2, \eta_3, \eta_4)\) in \(( -\infty, 0) \times ( -\infty, \infty) \times ( -\infty, 0) \times ( -\infty, \infty) \), satisfying the constraint \( \eta_1 \eta_4 = \eta_2 \eta_3 \) will in fact be the image of some \( \eta(\mu, \sigma_X^2, \sigma_Y^2) \), namely for \( \mu = -\eta_2/(2\eta_1), \sigma_X^2 = -1/(2\eta_1), \) and \( \sigma_Y^2 = -1/(2\eta_3) \). So, we have the space of values of the natural parameter is

\[ \eta(\Theta) = \{ (\eta_1, \eta_2, \eta_3, \eta_4) : \eta_1 < 0, \eta_3 < 0, \eta_1 \eta_4 = \eta_2 \eta_3 \}. \]

Now, we just have to see if we can find \( \theta_0, \ldots, \theta_4 \) as in the Proposition 2.5. to show that \( T \) is minimal sufficient. Let’s try the following

\[ \begin{align*}
\theta_0 &= (0, 1/2, 1/2) \\
\theta_1 &= (0, 1/4, 1/2) \\
\theta_2 &= (0, 1/2, 1/4) \\
\theta_3 &= (1/2, 1/2, 1/2) \\
\theta_4 &= (1/2, 1/4, 1/2),
\end{align*} \]

which gives

\[ \begin{bmatrix} \eta(\theta_0) \\ \eta(\theta_1) \\ \eta(\theta_2) \\ \eta(\theta_3) \\ \eta(\theta_4) \end{bmatrix} = \begin{bmatrix} -1 & 0 & -1 & 0 \\ -2 & 0 & -1 & 0 \\ -1 & 0 & -2 & 0 \\ -1 & 1 & -1 & 1 \\ -2 & 2 & -1 & 1 \end{bmatrix} \]
By direct calculation, the determinant of this $4 \times 4$ matrix is 1, so we have the desired linear independence. Also, by the same argument as above, the family is identifiable, and hence $T$ is minimal sufficient.

Note: $T$ is not complete in this case since

$$E_\theta \left[ n^{-1}T_2 - m^{-1}T_4 \right] = \mu - \mu = 0, \quad \forall \theta.$$
Now we consider the loss function. This is very problematic because most problems don’t come with a loss function and we generally choose something for mathematical convenience like squared error loss for estimation or 0-1 loss for hypothesis testing (lose 1 for a type II error, lose 0 for everything else). I have never had a client who had a clear loss function.

Then there is the typical situation where we restrict the allowable decision rules, e.g. to unbiased estimators or level $\alpha$ tests. Again, this is purely for mathematical convenience or because it is the custom. In many problems, there are real losses (or gains) from different decisions, but these are almost always very complicated and typically unknown. In the soil remediation example, there would be the loss of a fine for the company if the governmental regulators decided the company did not do a satisfactory job, which may be known, but even if they passed the regulations, there could be future law suits that could cost them a lot of money, and that isn’t known.

Furthermore, in practice, especially for big data sets, we need to restrict to “computationally feasible” decision rules, and this tends to be a class of decision rules which is not really tractable. Notice it is not even mentioned in the notes.

And finally, there is the objective of minimizing the risk. The risk must be computed using the model - again, that model may not be valid. Also, we tend to look at all parameter values in frequentist settings, and why should one worry about parameter values that seem unlikely, given the data. At least the Bayesian approach overcomes this criticism. The risk is the expected or average loss, but why should this be considered the only way to evaluate a decision rule?