**Problem Set 12.1, page 544**

1. When 7 is added to every output, the mean increases by 7 and the variance does not change (because new variance comes from \((\text{distance})^2\) to the new mean).

   New sample mean and new expected mean: Add 7. New variance: No change.

2. If we add \(\frac{1}{3}\) to \(\frac{1}{7}\) (fraction of integers divisible by 3 plus fraction divisible by 7) we have **double counted** the integers divisible by both 3 and 7. This is a fraction \(\frac{1}{21}\) of all integers (because these double counted numbers are multiples of 21). So the fraction divisible by 3 or 7 or both is:

   \[
   \frac{1}{3} + \frac{1}{7} - \frac{1}{21} = \frac{7}{21} + \frac{3}{21} - \frac{1}{21} = \frac{9}{21} = \frac{3}{7}
   \]

3. In the numbers from 1 to 1000, each group of ten numbers will contain each possible ending \(x = 1, 2, 3, \ldots, 0\). So those endings all have the same probability \(p_i = \frac{1}{10}\).

   Expected mean of that last digit \(x\):

   \[
   m = E[x] = \sum p_i x_i = \frac{1}{10} \sum_{i=0}^{9} i = \frac{45}{10} = 4.5
   \]

   The best way to find the variance \(\sigma^2 = 8.25\) is **in the last line below and in problem 12.1.7**. The slower way to find \(\sigma^2\) is:

   \[
   \sigma^2 = E[(x - 4.5)^2] = \sum_{i=0}^{9} p_i (x_i - 4.5)^2 = \frac{1}{10} \sum_{i=0}^{9} (i - 4.5)^2
   \]

   We can separate \((i - 4.5)^2\) into \((i^2 - 9i + (4.5)^2)\) and add from \(i = 0\) to \(i = 9\):

   \[
   \frac{1}{10} \left( \sum_{i=0}^{9} i^2 - 9 \sum_{i=0}^{9} i + \sum_{i=0}^{9} (4.5)^2 \right) = \frac{1}{10} \left( 285 - 9(45) + 10(4.5)^2 \right)
   \]

   \[
   = \frac{1}{10} \left( 285 - 405 + 202.5 \right) = \frac{82.5}{10} = 8.25 = \frac{33}{4}
   \]

   Notice that 202.5 is half of 405—like \(N m^2\) and \(2N m^2\) in equation (4), page 536.

   **I should have extended equation (4) to its best form**:

   \[
   \sigma^2 = E[(x - m)^2] = E[x^2] - m^2
   \]

   That quickly gives \(\frac{285}{10} - (4.5)^2 = 8.25 = \text{same answer}\.)
4 For numbers ending in 0, 1, 2, \ldots, 9 the squares end in \( x = 0, 1, 4, 9, 6, 5, 6, 9, 4, 1 \). So the probabilities of \( x = 0 \) and 5 are \( p = \frac{1}{10} \) and the probabilities of \( x = 1, 4, 6, 9 \) are \( p = \frac{1}{5} \). The mean is

\[
m = \sum p_i x_i = 0 \cdot \frac{0}{10} + \frac{5}{10} + \frac{1}{5} (1 + 4 + 6 + 9) = 4.5 = \text{same as before.}
\]

The variance using the improvement of equation (4) is

\[
\sigma^2 = E[x^2] - m^2 = \frac{1}{10} 0^2 + \frac{1}{10} 1^2 + \frac{1}{5} (1^2 + 4^2 + 6^2 + 9^2) - m^2
\]

\[
= \frac{25}{10} + \frac{134}{5} - 20.25 = 9.05
\]

5 For numbers from 1 to 1000, the first digit is \( x = 1 \) for 1000 and 100-199 and 10-19 and 1 (112 times). The first digit is \( x = 2 \) for 200-299 and 20-29 and 2 (111 times). The other first digits \( x = 3 \) to 9 also happen (111 times). Total (1000 times) is correct.

The average first digit is the mean, close to \( \frac{1}{9} (1 + 2 + \cdots + 9) = 5 \):

\[
m = \frac{\sum p_i x_i}{1000} = \frac{112}{1000} (1) + \frac{111}{1000} (2+3+\cdots+9) = \frac{112 + 111(44)}{1000} = \frac{4996}{1000} = 4.996 \approx 5.
\]

The variance is

\[
\sigma^2 = E[ (x - m)^2 ] = E[ x^2 ] - m^2 = \frac{112}{1000} (1^2) + \frac{111}{1000} (2^2 + \cdots + 9^2) - m^2
\]

\[
= \frac{112 + 111(284)}{1000} - m^2 \approx \frac{31635}{1000} - 5^2 = 6.635.
\]

6 The first digits of 157^2, 312^2, 696^2, and 602^2 are 2, 9, 4, 3. The sample mean is \( \frac{1}{4} (2 + 9 + 4 + 3) = \frac{18}{4} = 4.5 \). The sample variance with \( N - 1 = 3 \) is

\[
S^2 = \frac{1}{3} [ (-2.5)^2 + (4.5)^2 + (-.5)^2 + (-1.5)^2 ] = \frac{1}{3} \left[ \frac{29}{4} \right].
\]

7 This question is about the fast way to compute \( \sigma^2 \) using \( m^2 \). The mean \( m \) is probably already computed:

\[
\sigma^2 = \sum p_i (x_i - m)^2 = \sum p_i (x_i^2 - 2mx_i + m^2)
\]

\[
= \sum p_i x_i^2 - 2m \sum p_i x_i + m^2 \sum p_i
\]

\[
= \sum p_i x_i^2 - 2m^2 + m^2 = \sum p_i x_i^2 - m^2 = E[x^2] - m^2.
\]
For $N = 24$ samples, all equal to $x = 20$, 

$$\mu = \frac{1}{N} \sum x_i = \frac{24}{24}(20) = 20 \quad \text{and} \quad S^2 = \frac{1}{N-1} \sum (x_i - \mu)^2 = 0$$

For 12 samples of $x = 20$ and 12 samples of $x = 21$,

$$\mu = \frac{12(20) + 12(21)}{24} = 20.5 \quad \text{and} \quad S^2 = \frac{1}{N-1} \sum (x_i - \mu)^2 = \frac{1}{23} 24 \left(\frac{1}{2}\right)^2 = \frac{6}{23}$$

9 This question asks you to set up a random 0-1 generator and run it a million times to find the average $A_{1000000}$.

One way is to use MATLAB’s `rand` command with a uniform distribution between 0 and 1. Add $\frac{1}{2}$ to go between 0.5 and 1, then find the integer part (0 or 1). Using your computed average $A_N$ (its mean is $m = \frac{1}{2}$ since 0 and 1 are equally likely for every sample) find the normalized variable $X$:

$$X = \frac{A_N - \frac{1}{2}}{2\sqrt{N}} = \frac{A_N - \frac{1}{2}}{2000} \quad \text{for} \quad N = \text{one million}.$$ 

10 The average number of heads in $N$ fair coin flips is $m = N/2$. This is obvious—but how to derive it from probabilities $p_0$ to $p_N$ of 0 to $N$ heads? We have to compute

$$m = 0p_0 + 1p_1 + \cdots + Np_N \quad \text{with} \quad p_i = \frac{b_i}{2^N} = \frac{1}{2^N} \frac{N!}{i!(N-i)!}$$

A useful fact is $p_i = p_{N-i}$. The probability of $i$ heads equals the probability of $i$ tails. If we take just those two terms in $m$, they give

$$ip_i + (N-i)p_{N-i} = ip_i + (N-i)p_i = Np_i$$

So we can compute $m$ two ways and add:

$$m = 0p_0 + 1p_1 + \cdots + (N-1)p_{N-1} + Np_N$$

$$m = Np_0 + (N-1)p_1 + \cdots + 1p_{N-1} + 0p_0$$

$$2m = Np_0 + Np_1 + \cdots + Np_{N-1} + Np_N$$

$$= N(p_0 + p_1 + \cdots + p_{N-1} + p_N) = N.$$ 

Then $m = N/2$. The average number of heads is $N/2$. 

---

**Solutions to Exercises**

8 For $N = 24$ samples, all equal to $x = 20$,

$$\mu = \frac{1}{N} \sum x_i = \frac{24}{24}(20) = 20 \quad \text{and} \quad S^2 = \frac{1}{N-1} \sum (x_i - \mu)^2 = 0$$

For 12 samples of $x = 20$ and 12 samples of $x = 21$,

$$\mu = \frac{12(20) + 12(21)}{24} = 20.5 \quad \text{and} \quad S^2 = \frac{1}{N-1} \sum (x_i - \mu)^2 = \frac{1}{23} 24 \left(\frac{1}{2}\right)^2 = \frac{6}{23}$$

9 This question asks you to set up a random 0-1 generator and run it a million times to find the average $A_{1000000}$.

One way is to use MATLAB’s `rand` command with a uniform distribution between 0 and 1. Add $\frac{1}{2}$ to go between 0.5 and 1, then find the integer part (0 or 1). Using your computed average $A_N$ (its mean is $m = \frac{1}{2}$ since 0 and 1 are equally likely for every sample) find the normalized variable $X$:

$$X = \frac{A_N - \frac{1}{2}}{2\sqrt{N}} = \frac{A_N - \frac{1}{2}}{2000} \quad \text{for} \quad N = \text{one million}.$$ 

10 The average number of heads in $N$ fair coin flips is $m = N/2$. This is obvious—but how to derive it from probabilities $p_0$ to $p_N$ of 0 to $N$ heads? We have to compute

$$m = 0p_0 + 1p_1 + \cdots + Np_N \quad \text{with} \quad p_i = \frac{b_i}{2^N} = \frac{1}{2^N} \frac{N!}{i!(N-i)!}$$

A useful fact is $p_i = p_{N-i}$. The probability of $i$ heads equals the probability of $i$ tails. If we take just those two terms in $m$, they give

$$ip_i + (N-i)p_{N-i} = ip_i + (N-i)p_i = Np_i$$

So we can compute $m$ two ways and add:

$$m = 0p_0 + 1p_1 + \cdots + (N-1)p_{N-1} + Np_N$$

$$m = Np_0 + (N-1)p_1 + \cdots + 1p_{N-1} + 0p_0$$

$$2m = Np_0 + Np_1 + \cdots + Np_{N-1} + Np_N$$

$$= N(p_0 + p_1 + \cdots + p_{N-1} + p_N) = N.$$ 

Then $m = N/2$. The average number of heads is $N/2$. 

11 \( \mathbf{E}[x^2] = \mathbf{E}[(x - m)^2 + 2xm - m^2] \)
\[ = \mathbf{E}[(x - m)^2] + 2m \mathbf{E}[x] - m^2 \mathbf{E}[1] \]
\[ = \sigma^2 + 2m^2 - m^2 = \sigma^2 + m^2 \]

12 The first step multiplies two independent 1-dimensional integrals (each one from \(-\infty\) to \(\infty\)) to produce a 2-dimensional integral over the whole plane:
\[
2\pi \int_{-\infty}^{\infty} p(x) \, dx \int_{-\infty}^{\infty} p(y) \, dy = 2\pi \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x) p(y) \, dx \, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-x^2/2} e^{-y^2/2} \, dx \, dy.
\]
The second step changes to polar coordinates \((x = r \cos \theta, y = r \sin \theta, dx \, dy = r \, dr \, d\theta)\), \(x^2 + y^2 = r^2\) with \(0 \leq \theta \leq 2\pi\) and \(0 \leq r \leq \infty\). Notice \(-x^2/2 - y^2/2 = -r^2/2\):
\[
\int_{\text{plane}} \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} e^{-r^2/2} r \, dr \, d\theta = \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} e^{-r^2/2} r \, dr \, d\theta
\]
The \(r\) and \(\theta\) integrals give the answers 1 and \(2\pi\):
\[
\int_{r=0}^{\infty} e^{-r^2/2} r \, dr = \left[-e^{-r^2/2}\right]_{r=0}^{\infty} = 1 \quad \int_{\theta=0}^{2\pi} 1 \, d\theta = 2\pi.
\]
The trick was to get \(e^{-r^2/2} r \, dr\) (which is a perfect derivative of \(-e^{-r^2/2}\)) by combining \(e^{-x^2/2} \, dx\) and \(e^{-y^2/2} \, dy\) (which can not be separately integrated from \(a\) to \(b\)).

**Problem Set 12.2, page 554**

1 (a) Mean \(m = \mathbf{E}[x] = (0)(1 - p) + (1)(p) = p\) when the probability of heads is \(p\). Here \(x = 0\) for tails and \(x = 1\) for heads. Notice that \(0^2 = 0\) and \(1^2 = 1\) so \(\mathbf{E}[x^2] = \mathbf{E}[x] = p\).

\[ \text{Variance } \sigma^2 = \mathbf{E}[x^2] - m^2 = p - p^2 \]

(b) These are independent flips! So the \(N\) by \(N\) covariance matrix \(V\) is diagonal. The diagonal entries are the variances \(\sigma^2 = p - p^2\) for each flip. Then the rule \((16 - 17 - 18)\) gives the overall variance of the sum from \(N\) flips:
overall variance = \begin{bmatrix} 1 & 1 & \ldots & 1 \end{bmatrix}V \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = N\sigma^2 = N(p - p^2)

This is just saying: Add the variances for the \( N \) independent experiments. Here those \( N \) experiments just repeat one experiment.

2 I am just imitating equation (2) in the text. Now the experiments are numbered 3 and 5. They have means \( m_3 \) and \( m_5 \). The covariance \( \sigma_{35} \) adds up joint probabilities \( p_{ij} \) times (distance \( x_i - m_3 \)) times (distance \( y_j - m_5 \)). Here \( x_i \) and \( y_j \) are outputs from experiments 3 and 5:

\[
\sigma_{35} = \sum_{i} \sum_{j} p_{ij} (x_i - m_3) (y_j - m_5).
\]

3 The 3 by 3 covariance matrix \( V \) will be a sum of rank one matrices \( V_{ijk} = UU^T \) multiplied by the joint probability \( p_{ijk} \) of outputs \( x_i, y_j, z_k \). I am copying equation (12):

\[
V = \sum_{i} \sum_{j} \sum_{k} p_{ijk} UU^T
\]

These matrices \( UU^T = \) column times row are positive semidefinite with rank 1 (unless \( U = 0 \)). The sum \( V \) is positive definite unless the 3 experiments are dependent.

Notice that the means \( \overline{x}, \overline{y}, \overline{z} = m_1, m_2, m_3 \) have to be computed before the variances.

4 We are told that the 3 experiments are independent. Then the covariances are zero off the main diagonal of \( V \). This covariance matrix only shows “covariances with itself” = “variances” \( \sigma_1^2, \sigma_2^2, \sigma_3^2 \) on the main diagonal.

\[
V = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}.
\]
5 The point is that some output \( X = x_i \) must occur. So the possibilities are \( Y = y_j \) and \( X = x_1 \), or \( Y = y_j \) and \( X = x_2 \), or \( Y = y_j \) and \( X = x_3 \) et cetera. The total probability of \( Y = y_j \) is the sum of the conditional probabilities that \( Y = y_j \) when \( X = x_i \).

Here is another way to say this law of total probability. When \( B_1, B_2, \ldots \) are separate disjoint outcomes that together account for all possible outcomes, then for any \( A \)

\[
\text{Prob} (A) = \sum_i \text{Prob} (A \cap B_i) = \sum_i \text{Prob} (A|B_i) \text{Prob} (B_i).
\]

6 \( \text{Prob} (A|B) = \text{conditional probability} \) of \( A \) given \( B \) satisfies this axiom:

\[
\text{Prob} (A \text{ and } B) = \text{Prob} (A|B) \text{Prob} (B).
\]

Reason: If both \( A \) and \( B \) occur, then \( B \) must occur—and knowing that \( B \) occurs, \( \text{Prob} (A|B) \) gives the probability that \( A \) also occurs.

This axiom is saying that \( p_{ij} = \text{Prob} (A|B) p_i \)

where \( B \) is the event \( x = x_i \) which has \( \text{Prob} (B) = p_i \).

7 The joint probabilities \( p_{ij} = \text{Prob} (x = x_i \text{ and } y = y_j) \) are in the matrix \( P \):

\[
P = \begin{bmatrix}
0.1 & 0.3 \\
0.2 & 0.4
\end{bmatrix}
\]

with entries adding to 1.

Problem 6 says that \( \text{Prob} (Y = y_2 | X = x_1) = \frac{p_{12}}{p_{11} + p_{12}} = \frac{0.3}{0.1 + 0.3} = \frac{3}{4} \).

Problem 5 says that \( \text{Prob} (X = x_1) = p_{11} + p_{12} = 0.1 + 0.3 = 0.4 \).

8 This product rule of conditional probability is the axiom in Solution 12.2.6 above:

\[
\text{Prob} (A \text{ and } B) = \text{Prob} (A \text{ given } B) \text{ times } \text{Prob} (B).
\]
This discussion of Bayes’ Theorem is much too compressed! Let me reproduce three equations from Wolfram MathWorld. Here and are possible “sets” = “outcomes from an experiment” and the simple-looking identity (∗) connects conditional and unconditional probabilities.

We know from 8 that \[ \text{Prob}(A \text{ and } B) = \text{Prob}(A \text{ given } B) \times \text{Prob}(B) \]
Reversing and gives \[ \text{Prob}(A \text{ and } B) = \text{Prob}(B \text{ given } A) \times \text{Prob}(A) \]
(∗) Therefore \[ \text{Prob}(B \text{ given } A) = \frac{\text{Prob}(A \text{ given } B) \times \text{Prob}(B)}{\text{Prob}(A)} \]
MathWorld gives this extension to non-overlapping sets \(A_1, \ldots, A_n\) whose union is \(A\):
\[ \text{Prob}(A_i \text{ given } A) = \frac{\text{Prob}(A_i) \times \text{Prob}(A \text{ given } A_i)}{\sum_j \text{Prob}(A_j) \times \text{Prob}(A \text{ given } A_j)} \]

Problem Set 12.3, page 560

1. The two equations from two measurements are
\[ x = b_1 \]
\[ x = b_2 \]
or
\[ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} x \\ b_1 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \]
or \(Ax = b\).

The covariance matrix \(V\) is diagonal since the measurements are independent:
\[ V = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}. \]

The weighted least squares equation is \(A^T V^{-1} A \tilde{x} = A^T V^{-1} b\).

\[ A^T V^{-1} A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \]

\[ A^T V^{-1} b = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \frac{b_1}{\sigma_1^2} + \frac{b_2}{\sigma_2^2} \]

Then \(\tilde{x}\) is the ratio of those numbers:
\[ \tilde{x} = \frac{b_1/\sigma_1^2 + b_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2} \]
We are inverting a 5 \times 4 matrix using

\[
\begin{bmatrix}
\sigma_1^2 & \sigma_{12} \\
\sigma_{12} & \sigma_2^2
\end{bmatrix}^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} \begin{bmatrix}
\sigma_2^2 & -\sigma_{12} \\
-\sigma_{12} & \sigma_1^2
\end{bmatrix} = \rho \frac{\sigma_{12}}{\sigma_1 \sigma_2}.
\]

Continue Problem 3 to find variances \(\sigma_x^2\) and \(\sigma_y^2\) and to see covariance \(\sigma_{xy} = 0\):

\[
\begin{align*}
\int \int (x - m_x)^2 p(x, y) \, dx \, dy &= \int \int (x - m_x)^2 p(x) \, dx \int p(y) \, dy = \sigma_x^2 \\
\int \int (y - m_y) p(x, y) \, dx \, dy &= \int \int (y - m_y) p(x) \, dx \int p(y) \, dy = (0) (0).
\end{align*}
\]

We are inverting a 2 by 2 matrix using

\[
V^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 \sigma_1^2 \sigma_2^2 - \sigma_{12}^4} \begin{bmatrix}
\sigma_2^2 & -\sigma_{12} \\
-\sigma_{12} & \sigma_1^2
\end{bmatrix} = \rho \frac{1}{\sigma_1^2} \begin{bmatrix}
1/\sigma_1^2 & -\rho/\sigma_1 \sigma_2 \\
-\rho/\sigma_1 \sigma_2 & 1/\sigma_2^2
\end{bmatrix}.
\]
6 The right hand side of $\hat{x}_{k+1}$ shows the gain factor $1/(k + 1)$:

$$\hat{x}_k + \frac{1}{k+1}(b_{k+1} - \hat{x}_k) = \frac{b_1 + \cdots + b_k}{k} + \frac{1}{k+1} \left( b_{k+1} - \frac{b_1 + \cdots + b_k}{k} \right) = \frac{b_1 + \cdots + b_{k+1}}{k+1}$$

Check that each number $b_1, b_2, \ldots, b_k, b_{k+1}$ is correctly divided by $k + 1$:

$$\frac{1}{k} - \frac{1}{k+1} = \frac{1}{k} \left( \frac{k + 1 - 1}{k+1} \right) = \frac{1}{k+1}.$$ 

7 We are considering the case when all the measurements $b_1, b_2, \ldots, b_{k+1}$ have the same variance $\sigma^2$. We know that the correct variance of their average is $W_{k+1} = \sigma^2/(k+1)$.

We want to see how this answer comes from equation (18) when we have the correct $W_k = \sigma^2/k$ from the previous step, and we update to $W_{k+1}$:

$$\text{(18) says that } W_{k+1}^{-1} = W_k^{-1} + A_{k+1}^T A_{k+1}^{-1} = \frac{k}{\sigma^2} + [1/\sigma^2] [1] = \frac{k}{\sigma^2} + \frac{1}{\sigma^2} = \frac{k + 1}{\sigma^2}.$$ 

So $W_{k+1} = \sigma^2/(k+1)$ is correct at the new step (and forever by induction).

8 The three equations have variances $\sigma^2, s^2, \sigma^2$ and they have zero covariances. (This makes the step-by-step Kalman filter possible.) We can divide the equations by $\sigma, s, \sigma$ to produce unit variances (which lead to ordinary unweighted least squares). We are given $F = 1$:

$$\begin{bmatrix} 1/\sigma & 0 \\ -1/s & 1/s \\ 0 & 1/\sigma \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} b_0/\sigma \\ 0 \\ b_1/\sigma \end{bmatrix} \text{ is our } A\mathbf{x} = \mathbf{b}.$$ 

The normal equation (now unweighted) is $A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$:

$$\begin{bmatrix} 1/\sigma^2 + 1/s^2 & -1/s^2 \\ -1/s^2 & 1/\sigma^2 + 1/s^2 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} b_0/\sigma^2 \\ b_1/\sigma^2 \end{bmatrix}.$$ 

The determinant of this $A^T A$ is $\det = \frac{1}{\sigma^4} + \frac{2}{\sigma^2 s^2}$. The solution is

$$\hat{x}_1 = \frac{1}{\det} \left( \frac{b_0}{\sigma^4} + \frac{b_0}{\sigma^2 s^2} + \frac{b_1}{\sigma^2 s^2} \right), \quad \hat{x}_2 = \frac{1}{\det} \left( \frac{b_0}{\sigma^2 s^2} + \frac{b_1}{\sigma^2 s^2} + \frac{b_1}{\sigma^4} \right).$$
9 With $A = I$ and $u^T = v^T = [1 \ 2 \ 3]$ we can use the direct formula for $M^{-1}$:

$$(I - uv^T)^{-1} = I + \frac{uv^T}{1 - v^Tu} = I + \frac{1}{1 - 14} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 - \frac{1}{13} & \frac{2}{13} & \frac{3}{13} \\ \frac{2}{13} & 1 - \frac{4}{13} & \frac{6}{13} \\ \frac{3}{13} & \frac{6}{13} & 1 - \frac{9}{13} \end{bmatrix}.$$ Multiply $b = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$ to get $y = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} - \frac{16}{13} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \\ -19 \\ 4 \end{bmatrix}$.

Instead of this formula for $(I = u v^T)^{-1}$, try steps 1 and 2:

**Step 1** with $A = I$ gives $x = b$ and $z = u$.

**Step 2** gives $y = b - \frac{v^T u}{13} = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} - \frac{16}{13} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ as before.

10 We are asked to check that $M y = b$ using the update formula. Start with

$$M y = (A - uv^T) \left( x + \frac{v^T x}{c} z \right) = A x - u (v^T x) + \frac{v^T x A z}{c} - \frac{u (v^T z) (v^T x)}{c}$$

Since $Ax = b$ we hope the other 3 terms combine to give zero when $Az = u$

$$uv^T x \left[ -1 + \frac{1}{c} - \frac{v^T z}{c} \right] = \frac{uv^T x}{c} \left[ -c + 1 - v^T z \right] = 0$$

from the formula for $c$

11 Multiply columns times rows to see that the new $v$ changes $A^T A$ to

$$\begin{bmatrix} A^T & v \\ \hline v^T & 1 \end{bmatrix} = A^T A + vv^T$$

So adding the new row to $A$ (and of course the new column to $A^T$) has increased $A^T A$ by the rank one matrix $vv^T$. 
The book is ending with matrix multiplication! We could allow changes of rank \( r \):

When \( A \) changes to \( M = A − UW^{-1}V \), its inverse changes to

\[
M^{-1} = A^{-1} + A^{-1} U(W − VA^{-1}U)^{-1} VA^{-1}.
\]

This change has rank \( r \) when \( W_{r\times r} \) and \( V_{r\times n} \) and \( U_{n\times r} \) all have rank \( r \).