1. Linear estimator. Consider a channel with the observation $Y = XZ$, where the signal $X$ and the noise $Z$ are uncorrelated Gaussian random variables. Let $E[X] = 1$, $E[Z] = 2$, $\sigma^2_X = 5$, and $\sigma^2_Z = 8$.

(a) Find the MMSE linear estimate of $X$ given $Y$.
(b) Suppose your friend from Caltech tells you that he was able to derive an estimator with a lower MSE. Your friend from MIT disagrees, saying that this is not possible because the signal and the noise are Gaussian, and hence the MMSE linear estimator will also be the MMSE estimator. Could your MIT friend be wrong?

Solution:

(a) We know that the best linear estimate is given by the formula

$$\hat{X} = \frac{\text{Cov}(X,Y)}{\sigma^2_Y}(Y - E(Y)) + E(X).$$

Note that $X$ and $Z$ Gaussian and uncorrelated implies they are independent. Therefore,

$$E(Y) = E(XZ) = E(X)E(Z) = 2,$$
$$E(XY) = E(X^2Z) = E(X^2)E(Z) = (\sigma^2_X + E^2(X))E(Z) = 12,$$
$$E(Y^2) = E(X^2Z^2) = E(X^2)E(Z^2) = (\sigma^2_X + E^2(X)) (\sigma^2_Z + E^2(Z)) = 72,$$
$$\sigma^2_Y = E(Y^2) - E^2(Y) = 68,$$
$$\frac{\text{Cov}(X,Y)}{\sigma^2_Y} = \frac{E(XY) - E(X)E(Y)}{\sigma^2_Y} = \frac{5}{34}.$$ Using all of the above, we get

$$\hat{X} = \frac{5}{34}Y + \frac{12}{17}.$$ (b) The fact that the best linear estimate equals the best MMSE estimate when input and noise are independent Gaussians is only known to be true for additive channels. For multiplicative channels this need not be the case in general. In the following, we prove $Y$ is not Gaussian by contradiction. Suppose $Y$ is Gaussian, then $Y \sim N(2,68)$. We have

$$f_Y(y) = \frac{1}{\sqrt{2\pi \times 68}} e^{-\frac{(y-2)^2}{2 \times 68}}.$$
On the other hand, as a function of two random variables, $Y$ has pdf

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(x) f_Z\left(\frac{y}{x}\right) \, dx.$$  

But these two expressions are not consistent, because

$$f_Y(0) = \int_{-\infty}^{\infty} f_X(x) f_Z\left(0\right) \, dx = f_Z(0) \int_{-\infty}^{\infty} f_X(x) \, dx = f_Z(0)$$

$$= \frac{1}{\sqrt{2\pi} \times 8} e^{-\frac{(0)^2}{2 \times 8}}$$

$$\neq \frac{1}{\sqrt{2\pi} \times 68} e^{-\frac{(0-2)^2}{2 \times 68}} = f_Y(0),$$

which is a contradiction. Hence, $X$ and $Y$ are not joint Gaussian, and we might be able to derive an estimator with a lower MSE.

2. **Additive-noise channel with path gain.** Consider the additive noise channel shown in the figure below, where $X$ and $Z$ are zero mean and uncorrelated, and $a$ and $b$ are constants.

\[ X \xrightarrow{a} Z \xrightarrow{b} Y = b(aX + Z) \]

Find the MMSE linear estimate of $X$ given $Y$ and its MSE in terms only of $\sigma_X$, $\sigma_Z$, $a$, and $b$.

**Solution:** By the theorem of MMSE linear estimate, we have

$$\hat{X} = \frac{\text{Cov}(X, Y)}{\sigma_Y^2} (Y - \text{E}(Y)) + \text{E}(X).$$

Since $X$ and $Z$ are zero mean and uncorrelated, we have

$$\text{E}(X) = 0,$$

$$\text{E}(Y) = b(a\text{E}(X) + \text{E}(Z)) = 0,$$

$$\text{Cov}(X, Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y) = \text{E}(Xb(aX + Z)) = a b \sigma_X^2,$$

$$\sigma_Y^2 = \text{E}(Y^2) - (\text{E}(Y))^2 = \text{E}(b^2(aX + Z)^2) = b^2 a^2 \sigma_X^2 + b^2 \sigma_Z^2.$$ 

Hence, the best linear MSE estimate of $X$ given $Y$ is given by

$$\hat{X} = \frac{a \sigma_X^2}{b a^2 \sigma_X^2 + b \sigma_Z^2} Y.$$
3. **Image processing.** A pixel signal $X \sim U[-k, k]$ is digitized to obtain

$$\tilde{X} = i + \frac{1}{2}, \text{ if } i < X \leq i + 1, \quad i = -k, -k + 1, \ldots, k - 2, k - 1.$$ 

To improve the visual appearance, the digitized value $\tilde{X}$ is dithered by adding an independent noise $Z$ with mean $E(Z) = 0$ and variance $\text{Var}(Z) = N$ to obtain $Y = \tilde{X} + Z$.

(a) Find the correlation of $X$ and $Y$.
(b) Find the MMSE linear estimate of $X$ given $Y$. Your answer should be in terms only of $k$, $N$, and $Y$.

**Solution:**

(a) From the definition of $\tilde{X}$, we know $P\{\tilde{X} = i + \frac{1}{2}\} = P\{i < X \leq i + 1\} = \frac{1}{2k}$. By the law of total expectation, we have

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X(\tilde{X} + Z)) = E(X \tilde{X})$$

$$= \sum_{i=-k}^{k-1} E[X \tilde{X} | i < X \leq i + 1]P(i < X \leq i + 1)$$

$$= \sum_{i=-k}^{k-1} \int_i^{i+1} x(i + \frac{1}{2}) \frac{1}{2k} dx = \frac{1}{8k} \sum_{i=-k}^{k-1} (2i + 1)^2 = \frac{1}{4k} \sum_{i=1}^{k} (2i - 1)^2$$

$$= \frac{4k^2 - 1}{12}.$$

Since, $\sum_{i=1}^{k} i^2 = k(k + 1)(2k + 1)/6$.

(b) We have

$$E(X) = 0,$$

$$E(Y) = E(\tilde{X}) + E(Z) = 0,$$

$$\sigma_Y^2 = \text{Var}\tilde{X} + \text{Var}Z = \sum_{i=-k}^{k-1} (i + \frac{1}{2})^2 \frac{1}{2k} + N = \frac{1}{4k} \sum_{i=0}^{k-1} (2i + 1)^2 + N = \frac{4k^2 - 1}{12} + N.$$

Then, the best linear MMSE estimate of $X$ given $Y$ is given by

$$\hat{X} = \frac{\text{Cov}(X, Y)}{\sigma_Y^2} (Y - E(Y)) + E(X) = \frac{4k^2 - 1}{\frac{4k^2 - 1}{12} + N} Y$$

$$= \frac{4k^2 - 1}{4k^2 - 1 + 12N} Y.$$

4. **Noise cancellation.** A classical problem in statistical signal processing involves estimating a weak signal (e.g., the heart beat of a fetus) in the presence of a strong interference (the heart beat of its mother) by making two observations; one with the weak signal present and one without (by placing one microphone on the mother’s belly and another close to her heart).
The observations can then be combined to estimate the weak signal by “canceling out” the interference. The following is a simple version of this application.

Let the weak signal $X$ be a random variable with mean $\mu$ and variance $P$, and the observations be $Y_1 = X + Z_1$ ($Z_1$ being the strong interference), and $Y_2 = Z_1 + Z_2$ ($Z_2$ is a measurement noise), where $Z_1$ and $Z_2$ are zero mean with variances $N_1$ and $N_2$, respectively. Assume that $X$, $Z_1$ and $Z_2$ are uncorrelated. Find the MMSE linear estimate of $X$ given $Y_1$ and $Y_2$ and its MSE. Interpret the results.

**Solution:** This is a vector linear MSE problem. Since $Z_1$ and $Z_2$ are zero mean, $\mu_X = \mu_{Y_1} = \mu$ and $\mu_{Y_2} = 0$. We first normalize the random variables by subtracting off their means to get $X' = X - \mu$, and

$${\mathbf{Y'}} = \begin{bmatrix} Y_1 - \mu \\ Y_2 \end{bmatrix}.$$ 

Now using the orthogonality principle we can find the best linear MSE estimate $\hat{X}'$ of $X'$. To do so we first find

$$\Sigma_X = \begin{bmatrix} P + N_1 & N_1 \\ N_1 & N_1 + N_2 \end{bmatrix}$$

and

$$\Sigma_{XY} = \begin{bmatrix} P \\ 0 \end{bmatrix}.$$ 

Thus,

$$\hat{X}' = \Sigma_{XY}^{-1} \Sigma_X^{-1} \mathbf{Y'}$$

$$= \frac{P}{P(N_1 + N_2) + N_1N_2} \begin{bmatrix} N_1 + N_2 & -N_1 \\ -N_1 & P + N_1 \end{bmatrix} \begin{bmatrix} Y_1 - \mu \\ Y_2 \end{bmatrix}$$

$$= \frac{P}{P(N_1 + N_2) + N_1N_2} \begin{bmatrix} (N_1 + N_2)(Y_1 - \mu) - N_1Y_2 \\ P((N_1 + N_2)Y_1 - N_1Y_2) + N_1N_2\mu \end{bmatrix}.$$ 

The best linear MSE estimate is $\hat{X} = \hat{X}' + \mu$. Thus,

$$\hat{X} = \frac{P}{P(N_1 + N_2) + N_1N_2} \begin{bmatrix} (N_1 + N_2)(Y_1 - \mu) - N_1Y_2 \\ P((N_1 + N_2)Y_1 - N_1Y_2) + N_1N_2\mu \end{bmatrix}.$$ 

The MSE can be calculated by

$$\text{MSE} = \sigma_X^2 - \Sigma_{XY}^{-1} \Sigma_X^{-1} \Sigma_{XY}$$

$$= P - \frac{P}{P(N_1 + N_2) + N_1N_2} \begin{bmatrix} (N_1 + N_2) & -N_1 \\ 0 & 0 \end{bmatrix}$$

$$= P - \frac{P^2(N_1 + N_2)}{P(N_1 + N_2) + N_1N_2}$$

$$= \frac{PN_1N_2}{P(N_1 + N_2) + N_1N_2}.$$ 

The equation for the MSE makes perfect sense. First, note that if $N_1$ and $N_2$ are held constant but $P$ goes to infinity, the MSE tends to $\frac{N_1N_2}{N_1+N_2}$. Next, note that if both $N_1$ and $N_2$ go to
infinitely, the MSE goes to $\sigma^2$, i.e., the estimate becomes worthless. Finally, note that if either $N_1$ or $N_2$ goes to 0, the MSE also goes to 0. This is because the estimator will then use the measurement with zero noise variance and perfectly determine the signal $X$.

5. **Nonlinear estimator.** Consider a channel with the observation $Y = XZ$, where the signal $X$ and the noise $Z$ are uncorrelated Gaussian random variables. Let $E[X] = 1$, $E[Z] = 2$, $\sigma_X^2 = 5$, and $\sigma_Z^2 = 8$.

(a) Using the fact that $E(W^3) = \mu + 3\mu^2$ and $E(W^4) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$ for $W \sim N(\mu, \sigma^2)$, find the mean and covariance matrix of $[X \ Y \ Y^2]^T$.

(b) Find the MMSE linear estimate of $X$ given $Y$ and the corresponding MSE.

(c) Find the MMSE linear estimate of $X$ given $Y^2$ and the corresponding MSE.

(d) Find the MMSE linear estimate of $X$ given $Y$ and $Y^2$ and the corresponding MSE.

(e) Compare your answers in parts (b) through (d). Is the MMSE estimate of $X$ given $Y$ (namely, $E(X|Y)$) linear? This would answer Problem 1 in Homework Set #6.

**Solution:**

(a) Since $X$ and $Z$ are uncorrelated Gaussian random variables, they are independent. We have

\[
\begin{align*}
E(X^2) &= \sigma_X^2 + E^2(X) = 5 + 1 = 6, \\
E(X^3) &= 1 + 3 \times 1 \times 5 = 16, \\
E(X^4) &= 1 + 6 \times 1 \times 5 + 3 \times 25 = 106.
\end{align*}
\]

\[
\begin{align*}
E(Z^2) &= \sigma_Z^2 + E^2(Z) = 8 + 4 = 12, \\
E(Z^3) &= 2 + 3 \times 2 \times 8 = 50, \\
E(Z^4) &= 2^4 + 6 \times 4 \times 8 + 3 \times 64 = 400.
\end{align*}
\]

Since $X$ and $Z$ are independent, we have

\[
\begin{align*}
E(Y) &= E(XZ) = E(X)E(Z) = 2, \\
E(Y^2) &= E(X^2Z^2) = E(X^2)E(Z^2) = 6 \times 12 = 72, \\
E(Y^3) &= E(X^3Z^3) = E(X^3)E(Z^3) = 16 \times 50 = 800, \\
E(Y^4) &= E(X^4)E(Z^4) = 106 \times 400 = 42400.
\end{align*}
\]

Therefore, the mean of $[X \ Y \ Y^2]^T$ is $[1 \ 2 \ 72]^T$.

\[
\begin{align*}
\text{Var}(Y) &= E(Y^2) - E^2(Y) = 72 - 4 = 68, \\
\text{Var}(Y^2) &= E(Y^4) - E^2(Y^2) = 37216, \\
\text{Cov}(X, Y) &= E(XY) - E(X)E(Y) = E(X^2)E(Z) - E(X)E(Y) = 10, \\
\text{Cov}(X, Y^2) &= E(XY^2) - E(X)E(Y^2) = E(X^3)E(Z^2) - E(X)E(Y^2) = 120, \\
\text{Cov}(Y, Y^2) &= E(YY^2) - E(Y)E(Y^2) = E(X^3)E(Z^3) - E(Y)E(Y^2) = 656.
\end{align*}
\]
Therefore, the covariance matrix of $[X \ Y \ Y^2]^T$ is
\[
\begin{bmatrix}
5 & 10 & 120 \\
10 & 68 & 656 \\
120 & 656 & 37216
\end{bmatrix}.
\]

(b) The MMSE linear estimate of $X$ given $Y$ is
\[
\hat{X} = \frac{\text{Cov}(X,Y)}{\text{Var}(Y)}(Y - \mathbb{E}(Y)) + \mathbb{E}(X) = \frac{10}{68}(Y - 2) + 1 = \frac{5}{34}Y + \frac{24}{34},
\]
and its MSE is given by
\[
\text{MSE} = \text{Var}(X) - \frac{\text{Cov}^2(X,Y)}{\text{Var}(Y)} = 5 - \frac{100}{68} = 3.5294.
\]

(c) The MMSE linear estimate of $X$ given $Y^2$ is
\[
\hat{X} = \frac{\text{Cov}(X,Y^2)}{\text{Var}(Y^2)}(Y^2 - \mathbb{E}(Y^2)) + \mathbb{E}(X) = \frac{120}{37216}(Y^2 - 72) + 1 = \frac{15}{4652}Y^2 + \frac{893}{1163},
\]
and its MSE is given by
\[
\text{MSE} = \text{Var}(X) - \frac{\text{Cov}^2(X,Y^2)}{\text{Var}(Y^2)} = 5 - \frac{14400}{37216} = 4.6131.
\]

(d) We first normalize the random variables by subtracting off their means to get
\[
X' = X - \mathbb{E}(X) = X - 2,
\]
\[
Y' = Y - \mathbb{E}(Y) = Y - 2,
\]
\[
Y'^2 = Y^2 - \mathbb{E}(Y^2) = Y^2 - 72.
\]
Using the covariance matrix in part a, we have
\[
\Sigma_{[Y \ Y^2]^T X} = [10 \ 120]^T,
\]
\[
\Sigma_{[Y \ Y^2]^T} = \begin{bmatrix}
68 & 656 \\
656 & 37216
\end{bmatrix}.
\]
Therefore,
\[
\hat{X}' = \Sigma_{[Y \ Y^2]^T X}^{-1} \Sigma_{[Y \ Y^2]^T} \begin{bmatrix}
Y' \\
Y'^2
\end{bmatrix} = 0.1397Y' + 0.0008Y'^2,
\]
and hence
\[
\hat{X} = \hat{X}' + \mathbb{E}(X) = 0.1397(Y - 2) + 0.0008(Y^2 - 72) + 1 = 0.1397Y + 0.0008Y^2 + 0.663.
\]
The corresponding MSE is given by
\[
\text{MSE} = \text{Var}(X) - \Sigma_{[Y \ Y^2]^T X}^{-1} \Sigma_{[Y \ Y^2]^T} \Sigma_{[Y \ Y^2]^T X} = 3.5115.
\]
(e) MSE linear estimate of $X$ given $Y$ and $Y^2$ results in the minimum MSE among the three. Therefore, MSE linear estimate of $X$ given $Y$ does not have the minimum MSE and MMSE estimate of $X$ given $Y$ is not linear.

6. Minimum waiting time. Let $X_1, X_2, \ldots$ be i.i.d. exponentially distributed random variables with parameter $\lambda$, i.e. $f_{X_i}(x) = \lambda e^{-\lambda x}$, for $x \geq 0$.

(a) Does $Y_n = \min\{X_1, X_2, \ldots, X_n\}$ converge in probability as $n$ approaches infinity?

(b) If it converges what is the limit?

(c) What about $Z_n = nY_n$?

Solution:

(a) For any set of values $X_i$’s, the sequence $Y_n$ is monotonically decreasing in $n$. Since the random variables are non-negative, it is reasonable to guess that $Y_n$ converges to 0. Now $Y_n$ will converge in probability to 0 if and only if for any $\epsilon > 0$, $\lim_{n \to \infty} P\{|Y_n| > \epsilon\} = 0$.

$$P\{|Y_n| > \epsilon\} = P\{Y_n > \epsilon\} = P\{X_1 > \epsilon, X_2 > \epsilon, \ldots, X_n > \epsilon\} = P\{X_1 > \epsilon\}P\{X_2 > \epsilon\} \ldots P\{X_n > \epsilon\} = (1 - F_X(\epsilon))(1 - F_X(\epsilon)) \ldots (1 - F_X(\epsilon)) = (1 - e^{-\lambda \epsilon})^n = e^{-\lambda n \epsilon}.$$ 

As $n$ goes to infinity (for any finite $\epsilon > 0$) this converges to zero. Therefore $Y_n$ converges to 0 in probability.

(b) The limit to which $Y_n$ converges in probability is 0.

(c) Does $Z_n = nY_n$ converges to 0 in probability? No, it does not. In fact,

$$P\{|Z_n| > \epsilon\} = P\{nY_n > \epsilon\} = P\{Y_n > \frac{\epsilon}{n}\} = e^{-\lambda \frac{\epsilon}{n}} = e^{-\lambda \epsilon}$$

which does not depend on $n$. So $Z_n$ does not converge to 0 in probability. Note that the distribution of $Z_n$ is exponential with parameter $\lambda$, the same as the distribution of $X_i$.

$$F_{Z_n}(z) = P\{Z_n < z\} = 1 - e^{-\lambda z}.$$ 

In conclusion, if $X_i$’s are i.i.d. $\sim \exp(\lambda)$, then

$$Y_n = \min_{1 \leq i \leq n} \{X_i\} \sim \exp(n\lambda).$$
and 
\[ Z_n = nY_n \sim \exp(\lambda). \]
Thus
\[ P\{Y_n > \epsilon\} = e^{-\lambda \epsilon n} \to 0, \text{ so } Y_n \to 0 \text{ in probability,} \]
but
\[ P\{Z_n > \epsilon\} = e^{-\lambda \epsilon} \not\to 0, \text{ so } Z_n \not\to 0 \text{ in probability.} \]

7. *Roundoff errors.* The sum of a list of 100 real numbers is to be computed. Suppose that these numbers are rounded off to the nearest integer so that each number has an error that is uniformly distributed in the interval \((-0.5, 0.5).\) Use the central limit theorem to estimate the probability that the total error in the sum of the 100 numbers exceeds 6.

**Solution:** Errors are independent and uniformly distributed in \((-0.5, 0.5),\) i.e. \(e_i \sim \text{Unif}(-0.5, 0.5).\) Using the Central Limit Theorem, the total error \(e = \sum_{i=1}^{100} e_i\) can be approximated as a Gaussian random variable with mean
\[
E(e) = \sum_{i=1}^{100} E(e_i) = 0
\]
and variance
\[
\sigma_e^2 = \sum_{i=1}^{100} \text{Var}(e_i) = 100 \cdot \frac{1}{12} = 8.33.
\]
Therefore the probability of the given event is
\[
P\{e > 6\} = P\left\{\frac{e - E(e)}{\sigma_e} > \frac{6 - 0}{\sqrt{8.33}}\right\} \\
\approx Q\left(\frac{6}{\sqrt{8.33}}\right) = 0.0188.
\]
If the error is interpreted as an absolute difference, then
\[
P\{|e| > 6\} = 2 \cdot P\{e > 6\} \approx 0.0376.
\]

8. *Polya urn.* An urn initially has one red ball and one white ball. Let \(X_1\) denote the name of the first ball drawn from the urn. Replace that ball and one like it. Let \(X_2\) denote the name of the next ball drawn. Replace it and one like it. Continue, drawing and replacing.
(a) Argue that the probability of drawing $k$ reds followed by $n-k$ whites is
\[
\frac{1}{2} \cdot \frac{2}{3} \cdots \frac{k}{k+1} \cdot \frac{1}{(k+2)} \cdots \frac{(n-k)}{(n+1)} = \frac{k!(n-k)!}{(n+1)!} = \frac{1}{(n+1) \binom{n}{k}}.
\]
(b) Let $P_n$ be the proportion of red balls in the urn after the $n$th drawing. Argue that
\[
\Pr\{P_n = \frac{k}{n+2}\} = \frac{1}{n+1}, \quad k = 1, 2, \ldots, n+1.
\]
This shows that $P_n$ tends to a uniformly distributed random variable in distribution, i.e.,
\[
\lim_{n \to \infty} \Pr\{P_n \leq t\} \to t, \quad 0 \leq t \leq 1.
\]
(c) What can you say about the behavior of the proportion $P_n$ if you started initially with one red ball in the urn and two white balls? Specifically, what is the limiting distribution of $P_n$? Can you show $\Pr\{P_n = \frac{k}{n+3}\} = \frac{k}{n+2}$, for $k = 1, 2, \ldots, n+1$?

Solution:
(a) Let $r$ be the number of red balls, $w$ the number of white balls and $t$ the total number of balls $r + w$ in the urn. The picture of the balls in the urn for each drawing is given below. The ratio $r/t$ or $w/t$ gives the probability for the next drawing.

<table>
<thead>
<tr>
<th>Drawing</th>
<th>Balls in Urn</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>●</td>
<td>$r/t = 1/2$</td>
</tr>
<tr>
<td>2nd</td>
<td>●●</td>
<td>$r/t = 2/3$</td>
</tr>
<tr>
<td></td>
<td>●●●</td>
<td>$r/t = 3/4$</td>
</tr>
<tr>
<td></td>
<td>●●●●</td>
<td>$r/t = k/k+1$</td>
</tr>
<tr>
<td></td>
<td>●●●●●</td>
<td>$w/t = 1/k + 2$</td>
</tr>
<tr>
<td></td>
<td>●●●●●●</td>
<td>$w/t = 2/k + 3$</td>
</tr>
<tr>
<td></td>
<td>●●●●●●●</td>
<td>$w/t = 3/k + 4$</td>
</tr>
<tr>
<td></td>
<td>●●●●●●●●</td>
<td>$w/t = n - k/n + 1$</td>
</tr>
</tbody>
</table>

The probability of drawing $k$ reds followed by $n-k$ whites is
\[
\frac{1}{2} \cdot \frac{2}{3} \cdots \frac{k}{k+1} \cdot \frac{1}{(k+2)} \cdots \frac{(n-k)}{(n+1)} = \frac{k!(n-k)!}{(n+1)!} = \frac{1}{(n+1) \binom{n}{k}}.
\]
In fact, regardless of the ordering, the probability of drawing $k$ reds in $n$ draws is given by this expression as well. Note that after the $n$th drawing, there are $k+1$ red balls and $n - k + 1$ white balls for a total of $n + 2$ balls in the urn.

(b) After the $n$th drawing, where $k$ red balls have been drawn, the proportion $P_n$ of red balls is
\[
P_n = \frac{\text{number of red balls in the urn}}{\text{total number of balls in the urn}} = \frac{k+1}{n+2}.
\]
There are \( \binom{n}{k} \) orderings of outcomes with \( k \) red balls and \( n-k \) white balls, and each ordering has the same probability. In fact, in the expression of the probability of drawing \( k \) reds followed by \( n-k \) whites, there will be just permutations of the numerators in the case of a different sequence of drawings. Therefore

\[
\Pr \left\{ P_n = \frac{k + 1}{n + 2} \right\} = \binom{n}{k} \cdot \frac{1}{\binom{n}{k}(n+1)}
\]
\[
= \frac{1}{n+1} \quad \text{for } k = 0, \ldots, n.
\]

All the proportions are equally probable (the probability does not depend on \( k \)) and \( P_n \) tends to a uniformly distributed random variable \( \sim \text{Unif}[0,1] \) in distribution.

(c) If there are one red ball and two white balls to start with:

<table>
<thead>
<tr>
<th>Drawing</th>
<th>Red Balls</th>
<th>White Balls</th>
</tr>
</thead>
<tbody>
<tr>
<td>1\textsuperscript{st} drawing</td>
<td>o o •</td>
<td>r/t = 1/3</td>
</tr>
<tr>
<td>2\textsuperscript{nd} drawing</td>
<td>o o •</td>
<td>r/t = 2/4</td>
</tr>
<tr>
<td>(k - 1)\textsuperscript{th} drawing</td>
<td>o o • •</td>
<td>r/t = k/k + 2</td>
</tr>
<tr>
<td>k\textsuperscript{th} drawing</td>
<td>o o • • •</td>
<td>w/t = 2/k + 3</td>
</tr>
<tr>
<td>(k + 1)\textsuperscript{th} drawing</td>
<td>o o • • • •</td>
<td>w/t = 3/k + 4</td>
</tr>
<tr>
<td>(k + 2)\textsuperscript{th} drawing</td>
<td>o o o • • • •</td>
<td>w/t = 4/k + 5</td>
</tr>
<tr>
<td>(n - 1)\textsuperscript{th} drawing</td>
<td>o o o • • • • • •</td>
<td>w/t = n - k + 1/n + 2</td>
</tr>
<tr>
<td>n\textsuperscript{th} drawing</td>
<td>o o o • • • • • • • • • •</td>
<td></td>
</tr>
</tbody>
</table>

The probability of drawing \( k \) reds followed by \( n-k \) whites is

\[
\frac{1}{3} \cdot \frac{2}{4} \cdots \frac{k}{(k+2)} \cdot \frac{2}{(k+3)} \cdots \frac{(n-k+1)}{(n+2)} = \frac{2k!(n-k+1)!}{(n+2)!}
\]

After the \( n\textsuperscript{th} \) drawing,

\[
P_n = \frac{\text{number of red balls in the urn}}{\text{total number of balls in the urn}} = \frac{k+1}{n+3}.
\]

Similarly as in part (b),

\[
\Pr \left\{ P_n = \frac{k + 1}{n + 3} \right\} = \binom{n}{k} \cdot \frac{2k!(n-k+1)!}{(n+2)!}
\]
\[
= \frac{n!}{k!(n-k)!} \cdot \frac{2k!(n-k+1)!}{(n+2)!}
\]
\[
= \frac{2(n-k+1)}{(n+2)(n+1)}.
\]
The distribution is linearly decreasing in \( p = \frac{k+1}{n+3} \).

\[
\Pr\{P_n = p\} = \left(\frac{-2(n+3)}{(n+2)(n+1)}\right)p + \left(\frac{2}{n+1}\right).
\]

Therefore the limiting distribution of \( P_n \) as \( n \) goes to infinity is a triangular distribution with density \( f_{P_n}(p) = 2(1-p) \).

Note: the Polya urn for this problem generates a process that is identical to the following mixture of Bernoulli’s

\[
\Theta \sim f_\Theta(\theta) = 2(1-\theta), \quad 0 \leq \theta < 1.
\]

Let \( X_i \sim \text{Bernoulli}(\Theta) \), then \( P_n \) converges to \( \Theta \) as \( n \) tends to infinity. Considering the general case of \( \lambda_1 \) red balls and \( \lambda_2 \) white balls to start with, the limiting distribution is

\[
f_{\Theta}(\theta) = C(\lambda_1, \lambda_2) \theta^{\lambda_1-1}(1-\theta)^{\lambda_2-1}
\]

where \( C(\lambda_1, \lambda_2) \) is a function of \( \lambda_1 \) and \( \lambda_2 \).