1. (4.44) Theorem 4.5.6, with $a = b = 1$, serves as the base case of our inductive argument. Assume that the statement holds for $n > 1$. That is, assume

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{1\leq i<j\leq n} Cov(X_i, X_j)$$

Now for the $n+1$ case we have

$$\begin{aligned}
Var\left(\sum_{i=1}^{n+1} X_i\right) &= Var\left(\sum_{i=1}^{n} X_i + X_{n+1}\right)\\
&= Var\left(\sum_{i=1}^{n} X_i\right) + Var(X_{n+1}) + 2Cov\left(\sum_{i=1}^{n} X_i, X_{n+1}\right) && (4.5.6)\\
&= \sum_{i=1}^{n+1} Var(X_i) + 2\sum_{1\leq i<j\leq n} Cov(X_i, X_j) + 2Cov\left(\sum_{i=1}^{n} X_i, X_{n+1}\right) && (Assumption)
\end{aligned}$$

Now, first notice that we can rewrite the second term as

$$2\sum_{1\leq i<j\leq n} Cov(X_i, X_j) = 2\sum_{j=2}^{n}\sum_{i=1}^{j-1} Cov(X_i, X_j)$$

It should be clear that we simply need third term above proves the $j = n+1$ case in this sum. Consider the following.

$$\begin{aligned}
Cov\left(\sum_{i=1}^{n} X_i, X_{n+1}\right) &= \mathbb{E}\left(X_{n+1}\sum_{i=1}^{n} X_i\right) - \mathbb{E}\left(\sum_{i=1}^{n} X_i\right)\mathbb{E}\left(X_{n+1}\right)\\
&= \mathbb{E}\left(\sum_{i=1}^{n} X_{n+1}X_i\right) - \mathbb{E}\left(\sum_{i=1}^{n} X_i\right)\mathbb{E}\left(X_{n+1}\right)\\
&= \sum_{i=1}^{n}\mathbb{E}\left(X_{n+1}X_i\right) - \sum_{i=1}^{n}\mathbb{E}\left(X_i\right)\mathbb{E}\left(X_{n+1}\right)\\
&= \sum_{i=1}^{n}\Big[\mathbb{E}\left(X_{n+1}X_i\right) - \mathbb{E}\left(X_i\right)\mathbb{E}\left(X_{n+1}\right)\Big]\\
&= \sum_{i=1}^{n} Cov(X_i, X_{n+1})\\
&= \sum_{j=n+1}^{n+1}\sum_{i=1}^{j-1} Cov(X_i, X_j)
\end{aligned}$$

Using this, we see

$$2\sum_{1\leq i<j\leq n} Cov(X_i, X_j) + 2Cov\left(\sum_{i=1}^{n} X_i, X_{n+1}\right) = 2\sum_{j=2}^{n+1}\sum_{i=1}^{j-1} Cov(X_i, X_j) = 2\sum_{1\leq i<j\leq n+1} Cov(X_i, X_j)$$

Hence

$$Var\left(\sum_{i=1}^{n+1} X_i\right) = \sum_{i=1}^{n+1} Var(X_i) + 2\sum_{1\leq i<j\leq n+1} Cov(X_i, X_j)$$

2. (4.63) Let $X = \log Z$. Then $X = \exp(Z)$. Recall that $\exp(\cdot)$ is a convex function so by Jensen's Inequality, we have

$$\mathbb{E}(X) = \mathbb{E}(\exp(Z)) \geq \exp(\mathbb{E}(Z)) = \exp(0) = 1$$

Therefore, $\mathbb{E}(X) \geq 1$

3. (5.3) First note that $Y_i = 0$ with probability $P(X_i \leq \mu) = F_X(\mu)$ and $Y_i = 1$ with probability $P(X_i > \mu) = 1 - F_X(\mu)$. This holds for all $1 \leq i \leq n$ so assuming that we consider 1 as a "success" we have $Y_i \sim Bern(1 - F_X(\mu))$. Hence for $Z = \sum_{i=1}^{n} Y_i$ we have that $Z \sim Binom(n, 1 - F_X(\mu))$.

4. (a) For $0 < t < h$, the function e^{tx} is nondecreasing and nonegative on $(0, \infty)$. Therefore the event $X \geq a$ corresponds to the event $e^{tX} \geq e^{ta}$. Here $e^{tX} \geq 0$ and $e^{ta} > 0$ so using the Markov-Inequality, we have

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{1}{e^{ta}}\mathbb{E}(e^{tX}) = e^{-ta}M_X(t)$$

   (b) For $-h < t < 0$, the function $f(y) = e^{ty} \geq 0$ is monotone *decreasing* in y. This implies the event $X \leq a$ corresponds to $f(X) \geq f(a)$. Let $Y = f(X)$ and $c = f(a)$. Then Y is a nonegative random variable and $f(a) > 0$ is a positive constant. Therefore, we can use Markov's inequality. That is

$$P(X \leq a) = P(Y \geq c) \leq \frac{1}{c}\mathbb{E}(Y) = \frac{\mathbb{E}(e^{tX})}{e^{at}} = e^{-at}M_X(t)$$

5. (a) First note that for *any* random variable X, and $a > 0$, $P(X \geq a) \leq P(X^2 \geq a^2)$. Using this fact we see

$$P(X \geq a) \leq P(X^2 \geq a^2) = P\big(X^2 + \frac{\sigma^2}{a} \leq a + \frac{\sigma^2}{a}\big)$$

   Now, here we have $X^2 + \frac{\sigma^2}{a} \geq 0$ and $a + \frac{\sigma^2}{a} > 0$ so we can appyly Markov's Inequality to find

$$\begin{aligned}
P\big(X^2 + \frac{\sigma^2}{a} \leq a + \frac{\sigma^2}{a}\big) &\leq \frac{1}{(a + \sigma^2/a)^2}\mathbb{E}\big((X + \sigma^2/a)^2\big) \\
&= \frac{\mathbb{E}(X^2 + 2X\sigma^2/a + \sigma^4/a^2)}{(a^2 + \sigma^2/a)^2} \\
&= \frac{\mathbb{E}(X^2) + 2\sigma^2/a\mathbb{E}(X) + \sigma^4/a^2}{(a^2 + \sigma^2/a)^2} \\
&= \frac{\sigma^2 + \sigma^4/a^2}{(a^2 + \sigma^2)/a^2} \\
&= \frac{a^2\sigma^2 + \sigma^4}{(a^2 + \sigma^2)^2} \\
&= \frac{\sigma^2(a^2 + \sigma^2)}{(a^2 + \sigma^2)^2} \\
&= \frac{\sigma^2}{a^2 + \sigma^2}
\end{aligned}$$

(b) First notice that $P(X \geq a) = 1 - P(X < a) = 1 - P(-X \geq a)$. Now, $-a > 0$ so we can apply the inequality from part a to see

$$P(X \geq a) = 1 - P(-X \geq a) \geq 1 - \frac{\sigma^2}{\sigma^2 + a^2} = \frac{a^2}{\sigma^2 + a^2}$$

6. First, suppose that $\hat{Y}$ is in fact optimal. Then $\beta = \frac{Cov(X,Y)}{Var(X)}$. Then we have

$$\widehat{Y} = \mathbb{E}(Y) + \frac{Cov(X,Y)}{Var(X)}\Big[X - \mathbb{E}(X)\Big]$$

Now consider $Cov(X, W) = Cov(X, \hat{Y} - Y)$.

$$\begin{aligned}
Cov(X, \hat{Y} - Y) &= Cov(X, \hat{Y}) - Cov(X, Y) \\
&= Cov\Big[X, \mathbb{E}(Y) + \frac{Cov(X,Y)}{Var(X)}\big[X - \mathbb{E}(X)\big]\Big] - Cov(X, Y) \\
&= \frac{Cov(X,Y)}{Var(X)}Cov(X, X) - Cov(X, Y) \\
&= \frac{Cov(X,Y)}{Var(X)}Var(X) - Cov(X, Y) \\
&= 0
\end{aligned}$$

Note that the third equality used the fact that all terms in $\hat{Y}$ are constants expect the $\frac{Cov(X,Y)}{Var(X)}X$ term. Moreover, linearity of covariance was used repeatedly.

Now assume that $Cov(X, W) = 0$. This implies $Cov(X, \hat{Y}) - Cov(X, Y) = 0$ so $Cov(X, Y) = Cov(X, \hat{Y})$. Now, expanding the second term as we did above, we see that

$$Cov(X, Y) = Cov(X, \hat{Y}) = Cov\Big[X, \mathbb{E}(Y) + \beta\big[X - \mathbb{E}(X)\big]\Big] = \beta Var(X)$$

Solving for β gives $\beta = \frac{Cov(X,Y)}{Var(X)}$. This shows that we attain the optimal parameter for the MSE prediction problem and thus $Cov(X, W) = 0$ implies we attain the optimal MSE predictor.

7. (a) Recall that the best predictor of Y on X is given by $\mathbb{E}(Y|X)$. To find this value, we will derive the conditional distribution $f_{Y|X}(y|x)$.

$$\begin{aligned}
f_{Y|X}(y|x) &= \frac{\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\Big\{-\frac{1}{2(1-\rho^2)}\Big((\frac{x-\mu_X}{\sigma_X})^2 - 2\rho(\frac{x-\mu_X}{\sigma_X})(\frac{y-\mu_Y}{\sigma_Y}) + (\frac{y-\mu_Y}{\sigma_Y})^2\Big)\Big\}}{\frac{1}{\sqrt{2\pi}\sigma_X}\exp\{-\frac{(x-\mu_X)^2}{2\sigma_X^2}\}}\\
&= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}\exp\Big\{-\frac{1}{2(1-\rho^2)}\Big(\rho^2(\frac{x-\mu_X}{\sigma_X})^2 - 2\rho(\frac{x-\mu_X}{\sigma_X})(\frac{y-\mu_Y}{\sigma_Y}) + (\frac{y-\mu_Y}{\sigma_Y})^2\Big)\\
&= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}\exp\Big\{-\frac{1}{2(1-\rho^2)}\Big(\rho(\frac{x-\mu_X}{\sigma_X}) - (\frac{y-\mu_Y}{\sigma_Y})\Big)^2\Big\}\\
&= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}\exp\Big\{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\Big(\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X) - (y-\mu_Y)\Big)^2\Big\}\\
&= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}\exp\Big\{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\Big(y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\Big)^2\Big\}
\end{aligned}$$

We recognize this as density of a normal distribtion. Specifically

$$Y|X \sim N\left(\mu_y - \rho(\sigma_Y/\sigma_X)(x-\mu_X), \sigma_Y^2(1-\rho^2)\right)$$

Thus the best linear predictor is given by $\mathbb{E}(Y|X) = \mu_y - \rho(\sigma_Y/\sigma_X)(x-\mu_X)$.

(b) Recall that the MSE of a predictor is giveb by $MSE(\hat{Y}) = Var(\hat{Y}) + Bias(\hat{Y})^2$. In in our case $\hat{Y} = Y|X$. Notice that the bias is given by

$$E(E(Y|X) - Y) = E(E(Y|X)) - E(Y) = E(Y) - E(Y) = 0$$

Thus $MSE(Y|X) = Var(Y|X)$. We found that $Y|X \sim N\left(\mu_y - \rho(\sigma_Y/\sigma_X)(x-\mu_X), \sigma_Y^2(1-\rho^2)\right)$. Hence the MSE prediction error is given by

$$MSE(Y|X) = \sigma_Y^2\sqrt{1-\rho^2}$$

8. (a) Yes. Consider $X \sim N(0,1)$ and $Y = X^2$. Then

$$Cov(X,Y) = Cov(X,X^2) = \mathbb{E}(X^3) - \mathbb{E}(X^2)\mathbb{E}(X) = \mathbb{E}(X^3)$$

Then using the moment generating function of a standard normal, we see

$$\frac{\partial^3}{\partial t^3}\exp(1/2t^2)\Big|_{t=0} = t\exp(1/2t^2) + 2t\exp(1/2t^2) + t^3\exp(1/2t^2)\Big|_{t=0} = 0$$

Thus, we see that $f(X)$ and X are uncorrelated.

(b) No. Let $f(\cdot)$ be an arbitrary Borel-measurable function and let $Y = f(Y)$. Then

$$F_Y(y) = P(Y \leq y) = P(f(X) \leq y) = P(X \leq f^{-1}(y)) = F_X(f^{-1}(y))$$

So we see the distribution of Y is dependent on X. Now, given X we see

$$P(Y \leq y|X) = P(f(X) \leq y|X) = \begin{cases} 0 & f(X) \leq y \\ 1 & f(X) > y \end{cases}$$

So in the case that X is constant, Y is constant, but still relies on X. In the case that X is nonconstant, $F_X(f^{-1}(x))$ need not be constant. Thus $f(X)$ and X cannot be independent.

9. (a) Let $Y_1, Y_2 \overset{iid}{\sim} F_Y(y)$. Let $M = \max(Y_1, Y_2)$ and let m be the median of $F_Y(y)$. Then M is the largest order statistic and has cumulative distribution function $G_M(t) = (F_Y(t))^2$. Using this, we can calculate the desired probability

$$P(M > m) = 1 - P(M \leq m) = 1 - G_M(m) = 1 - (F_Y(m))^2 = 1 - (1/2)^2 = \frac{3}{4}$$

Here, the fourth equality used the fact that m was the median of $F_Y(y)$.

(b) Now, let $Y_1, Y_2, \dots, Y_n \sim F_Y(y)$, $M = \max(Y_i)_{i=1}^n$, and m be the median of $F_Y(y)$. Then M has CDF $G_M(t) = (F_y(t))^n$. With this, we can compute the desired probability.

$$P(M > m) = 1 - P(M \leq m) = 1 - G_M(m) = 1 - (F_Y(m))^n = 1 - (1/2)^n$$

10. Recall that if Y_k is the kth order statistic of $X_1, X_2, \dots, X_n$ then $U_k = F_X(Y_k)$ is the kth order statistic of a sample of size n from a Uniform on $[0,1]$. Now recall that the distribution of the kth order statistic is given by the following

$$\begin{aligned} g_{U_k}(y) &= \frac{n!}{(k-1)!(n-k)!} F_U(y)^{k-1}(1 - F_U(y))^{n-k} \\ &= \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} y^{k-1}(1-y)^{n-k} \end{aligned}$$

which we recognize as the Beta density function with parameters $(k, n-k+1)$. Hence, we can use the form $\mathbb{E}\big[F(Y_k)^2\big] = Var(F(Y_K)) + \mathbb{E}(F(Y_k))^2$ to find the desired value.

$$\begin{aligned} \mathbb{E}\big[F(Y_k)^2\big] &= \frac{k(n-k+1)}{(n+1)^2(n+2)} + \frac{k^2}{(n+1)^2} \\ &= \frac{k(n-k+1) + k^2(n+2)}{(n+1)^2(n+2)} \end{aligned}$$

11. First recall that $Z = F(Y_n)$ is the nth order statistic of the uniform on $(0,1)$. Thus, as $n \to \infty$ we expect $Z \to 1$.

$$P(|Z - 1| < \epsilon) = P(Z > 1 - \epsilon) = 1 - P(Z \leq 1 - \epsilon)$$

Since the cumulative distribution function of Z is given by $F_U(\cdot)^n$ where $F_U(\cdot)$ is the distribition function of the uniform on $(0,1)$ we have

$$1 - P(Z \leq 1 - \epsilon) = 1 - F_U(1-\epsilon)^n = 1 - (1-\epsilon)^n \to 1 \quad \text{as} \quad n \to \infty$$

Hence we see that Z converges in probability to 1. Specifically for $\epsilon = \frac{t}{n}$, we see that $P(Z \leq 1 - t/n) = (1 - t/n)^n \to e^{-t}$. Moreover,

$$P\big[n(1 - Z) \leq t\big] = P\big[1 - Z \leq t/n\big] = 1 - P\big[Z \leq 1 - t/n\big] \to 1 - e^{-t}$$

We recognize this as the CDF of an exponential distribution with rate parameter 1. Thus the limiting distribution of the nth order statistic on the unit interval is an exponential distribution.