Problem Set 8 Solutions

1. Let A_t (respectively, B_t) be a Bernoulli random variable that is equal to 1 if and only if the tth toss resulted in 1 (respectively, 2). We have $\mathbf{E}[A_tB_t] = 0$ (since $A_t \neq 0$ implies $B_t = 0$) and

$$\mathbf{E}[A_tB_s] = \mathbf{E}[A_t]\mathbf{E}[B_s] = \frac{1}{k} \cdot \frac{1}{k} \qquad \text{for} \quad s \neq t.$$

Thus,

$$\begin{aligned}\mathbf{E}[X_1X_2] &= \mathbf{E}\left[(A_1 + \cdots + A_n)(B_1 + \cdots + B_n)\right] \\ &= n\mathbf{E}\left[A_1(B_1 + \cdots + B_n)\right] \; = \; n(n-1) \cdot \frac{1}{k} \cdot \frac{1}{k}\end{aligned}$$

and

$$\begin{aligned}\text{cov}(X_1, X_2) &= \mathbf{E}[X_1X_2] - \mathbf{E}[X_1]E[X_2] \\ &= \frac{n(n-1)}{k^2} - \frac{n^2}{k^2} \; = \; -\frac{n}{k^2}.\end{aligned}$$

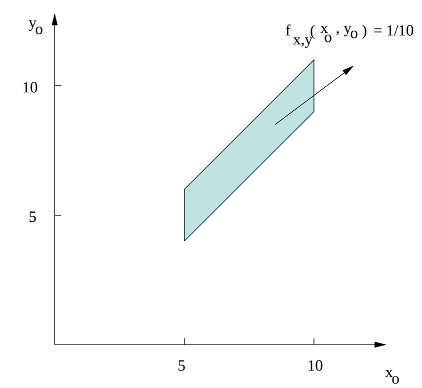
2. (a) The minimum mean squared error estimator $g(Y)$ is known to be $g(Y) = \mathbf{E}[X \mid Y]$. Let us first find $f_{X,Y}(x, y)$. Since $Y = X + W$, we can write

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{1}{2}, & \text{if } x - 1 \leq y \leq x + 1; \\ 0, & \text{otherwise} \end{cases}$$

and, therefore,

$$f_{X,Y}(x, y) = f_{Y|X}(y \mid x) \cdot f_X(x) = \begin{cases} \frac{1}{10}, & \text{if } x - 1 \leq y \leq x + 1 \text{ and } 5 \leq x \leq 10; \\ 0, & \text{otherwise} \end{cases}$$

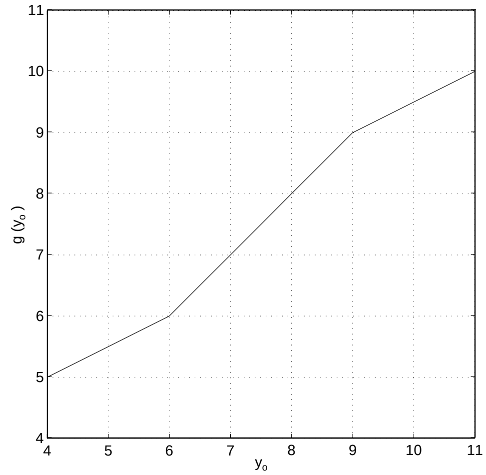
as shown in the plot below.



We now compute $\mathbf{E}[X \mid Y]$ by first determining $f_{X|Y}(x \mid y)$. This can be done by looking at the horizontal line crossing the compound PDF. Since $f_{X,Y}(x, y)$ is uniformly distributed in the defined region, $f_{X|Y}(x \mid y)$ is uniformly distributed as well. Therefore,

$$g(y) = \mathbf{E}[X \mid Y = y] = \begin{cases} \frac{5+(y+1)}{2}, & \text{if } 4 \leq y < 6; \\ y, & \text{if } 6 \leq y \leq 9; \\ \frac{10+(y-1)}{2}, & \text{if } 9 < y \leq 11. \end{cases}$$

The plot of $g(y)$ is shown here.



(b) The linear least squares estimator has the form

$$g_L(Y) = \mathbf{E}[X] + \frac{\text{cov}(X,Y)}{\sigma_Y^2}(Y - \mathbf{E}[Y]),$$

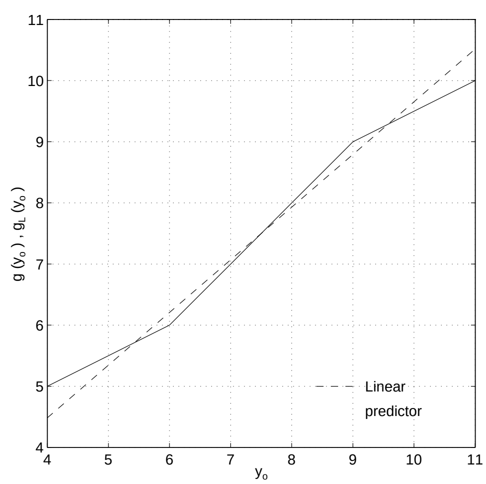
where $\text{cov}(X,Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$. We compute $\mathbf{E}[X] = 7.5$, $\mathbf{E}[Y] = \mathbf{E}[X] + \mathbf{E}[W] = 7.5$, $\sigma_X^2 = (10-5)^2/12 = 25/12$, $\sigma_W^2 = (1-(-1))^2/12 = 4/12$ and, using the fact that X and W are independent, $\sigma_Y^2 = \sigma_X^2 + \sigma_W^2 = 29/12$. Furthermore,

$$\begin{aligned}
\text{cov}(X,Y) &= \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\
&= \mathbf{E}[(X - \mathbf{E}[X])(X - \mathbf{E}[X] + W - \mathbf{E}[W])] \\
&= \mathbf{E}[(X - \mathbf{E}[X])(X - \mathbf{E}[X])] + \mathbf{E}[(X - \mathbf{E}[X])(W - \mathbf{E}[W])] \\
&= \sigma_X^2 + \mathbf{E}[(X - \mathbf{E}[X])]\mathbf{E}[(W - \mathbf{E}[W])] \;=\; \sigma_X^2 \;=\; 25/12.
\end{aligned}$$

Note that we use the fact that $(X - \mathbf{E}[X])$ and $(W - \mathbf{E}[W])$ are independent and $\mathbf{E}[(X - \mathbf{E}[X])] = 0 = \mathbf{E}[(W - \mathbf{E}[W])]$. Therefore,

$$g_L(Y) = 7.5 + \frac{25}{29}(Y - 7.5).$$

The linear estimator $g_L(Y)$ is compared with $g(Y)$ in the following figure. Note that $g(Y)$ is piecewise linear in this problem.



3. (a) The Chebyshev inequality yields $\mathbf{P}(|X-7| \geq 3) \leq \frac{9}{3^2} = 1$, which implies the uninformative/useless bound $\mathbf{P}(4 < X < 10) \geq 0$.

   (b) We will show that $\mathbf{P}(4 < X < 10)$ can be as small as 0 and can be arbitrarily close to 1. Consider a random variable that equals 4 with probability 1/2, and 10 with probability 1/2. This random variable has mean 7 and variance 9, and $\mathbf{P}(4 < X < 10) = 0$. Therefore, the lower bound from part (a) is the best possible.

   Let us now fix a small positive number ϵ and another positive number c, and consider a discrete random variable X with PMF

$$p_X(x) = \begin{cases} 0.5-\epsilon, & \text{if } x = 4+\epsilon; \\ 0.5-\epsilon, & \text{if } x = 10-\epsilon; \\ \epsilon, & \text{if } x = 7-c; \\ \epsilon, & \text{if } x = 7+c. \end{cases}$$

   This random variable has a mean of 7. Its variance is

$$(0.5-\epsilon)(3-\epsilon)^2 + (0.5-\epsilon)(3-\epsilon)^2 + 2\epsilon c^2$$

   and can be made equal to 9 by suitably choosing c. For this random variable, we have $\mathbf{P}(4 < X < 10) = 1 - 2\epsilon$, which can be made arbitrarily close to 1.

   On the other hand, this probability can not be made equal to 1. Indeed, if this probability were equal to 1, then we would have $|X-7| \leq 3$, which would imply that the variance in less than 9.

4. Consider a random variable X with PMF

$$p_X(x) = \begin{cases} p, & \text{if } x = \mu - c; \\ p, & \text{if } x = \mu + c; \\ 1-2p, & \text{if } x = \mu. \end{cases}$$

   The mean of X is μ, and the variance of X is $2pc^2$. To make the variance equal σ^2, set $p = \frac{\sigma^2}{2c^2}$. For this random variable, we have

$$\mathbf{P}(|X-\mu| \geq c) = 2p = \frac{\sigma^2}{c^2},$$

   and therefore the Chebyshev inequality is tight.

5. Note that n is deterministic and H is a random variable.

   (a) Use $X_1, X_2, \ldots$ to denote the (random) measured heights.

$$\begin{aligned} H &= \frac{X_1 + X_2 + \cdots + X_n}{n} \\ \mathbf{E}[H] &= \frac{\mathbf{E}[X_1 + X_2 + \cdots + X_n]}{n} = \frac{n\mathbf{E}[X]}{n} = h \\ \sigma_H &= \sqrt{\text{var}(H)} = \sqrt{\frac{n\,\text{var}(X)}{n^2}} \quad \text{(var of sum of \textit{independent} r.v.s is sum of vars)} \\ &= \frac{1.5}{\sqrt{n}} \end{aligned}$$

(b) We solve $\frac{1.5}{\sqrt{n}} < 0.01$ for n to obtain $n > 22500$.

(c) Apply the Chebyshev inequality to H with $\mathbf{E}[H]$ and $\text{var}(H)$ from part (a):

$$\begin{aligned} \mathbf{P}(|H-h| \geq t) &\leq \left(\frac{\sigma_H}{t}\right)^2 \\ \mathbf{P}(|H-h| < t) &\geq 1-\left(\frac{\sigma_H}{t}\right)^2 \end{aligned}$$

To be "99% sure" we require the latter probability to be at least 0.99. Thus we solve

$$1-\left(\frac{\sigma_H}{t}\right)^2 \geq 0.99$$

with $t = 0.05$ and $\sigma_H = \frac{1.5}{\sqrt{n}}$ to obtain

$$n \geq \left(\frac{1.5}{0.05}\right)^2 \frac{1}{0.01} = 90000.$$

(d) The variance of a random variable increases as its distribution becomes more spread out. In particular, if a random variable is known to be limited to a particular closed interval, the variance is maximized by having 0.5 probability of taking on each endpoint value. In this problem, random variable X has an unknown distribution over $[0, 3]$. The variance of X cannot be more than the variance of a random variable that equals 0 with probability 0.5 and 3 with probability 0.5. This translates to the standard deviation not exceeding 1.5.
In fact, this argument can be made more rigorous as follows.
First, we have

$$\text{var}(X) \leq \mathbf{E}[(X-\frac{3}{2})^2] = \mathbf{E}[X^2] - 3\mathbf{E}[X] + \frac{9}{4}$$

since $\mathbf{E}[(X-a)^2]$ is minimized when a is the mean (i.e., the mean is the least-squared estimator).
Second, we also have

$$0 \leq \mathbf{E}[X(3-X)] = \mathbf{E}[X] - \mathbf{E}[X^2]$$

since the variable has support in $[0, 3]$. Adding the above two inequalities, we have

$$\text{var}(X) \leq \frac{9}{4}$$

or equivalently, $\sigma_X \leq \frac{3}{2}$.

6. First, let's calculate the expectation and the variance for Y_n, T_n, and A_n.

$$\begin{aligned} Y_n &= (0.5)^n X_n \\ T_n &= Y_1 + Y_2 + \cdots + Y_n \\ A_n &= \frac{1}{n} T_n \end{aligned}$$

$$\begin{aligned}
\mathbf{E}[Y_n] &= \mathbf{E}\left[(\tfrac{1}{2})^n X_n\right] = (\frac{1}{2})^n \mathbf{E}[X_n] = \mathbf{E}[X](\tfrac{1}{2})^n = 2(\tfrac{1}{2})^n \\
\mathrm{var}(Y_n) &= \mathrm{var}\left((\tfrac{1}{2})^n X_n\right) = (\tfrac{1}{2})^{2n}\mathrm{var}(X_n) = \mathrm{var}(X)(\tfrac{1}{2})^{2n} = 9(\tfrac{1}{4})^{2n} \\
\mathbf{E}[T_n] &= \mathbf{E}[Y_1 + Y_2 + \cdots + Y_n] = \mathbf{E}[Y_1] + \mathbf{E}[Y_2] + \cdots + \mathbf{E}[Y_n] \\
&= 2\sum(\tfrac{1}{2})^i = 2\frac{0.5(1-0.5^n)}{1-0.5} = 2\left(1-(\tfrac{1}{2})^n\right) \\
\mathrm{var}(T_n) &= \mathrm{var}(Y_1 + Y_2 + \cdots + Y_n) = \sum_{i=1}^{n}(\tfrac{1}{4})^i \mathrm{var}(X_i) \\
&= 9\left(\frac{\frac{1}{4}\left(1-(\frac{1}{4})^n\right)}{1-\frac{1}{4}}\right) = 3\left(1-(\tfrac{1}{4})^n\right) \\
\mathbf{E}[A_n] &= \mathbf{E}\left[\frac{1}{n}T_n\right] = \frac{1}{n}\mathbf{E}[T_n] = \frac{2}{n}\left(1-(\tfrac{1}{2})^n\right) \\
\mathrm{var}(A_n) &= \mathrm{var}\left(\frac{1}{n}T_n\right) = \left(\frac{1}{n}\right)^2 \mathrm{var}(T_n) = \frac{3}{n^2}\left(1-\left(\frac{1}{4}\right)^n\right)
\end{aligned}$$

(a) Yes. Y_n converges to 0 in probability. As n becomes very large, the expected value of Y_n approaches 0 and the variance of Y_n approaches 0. So, by the Chebychev Inequality, Y_n converges to 0 in probability.

(b) No. Assume that T_n converges in probability to some value a. We also know that:

$$\begin{aligned}
T_n &= Y_1 + (Y_2 + Y_3 + ...Y_n) \\
&= Y_1 + ((0.5)^2 X_2 + (0.5)^3 X_3 + \cdots + (0.5)^n X_n) \\
&= Y_1 + \frac{1}{2}(0.5X_2 + (0.5)^2 X_3 + \cdots + (0.5)^{n-1} X_n).
\end{aligned}$$

Notice that $0.5X_2 + (0.5)^2 X_3 + \cdots + (0.5)^{n-1} X_n$ converges to the same limit as T_n when n goes to infinity. If T_n is to converge to a, Y_1 must converge to $a/2$. But this is clearly false, which presents a contradiction in our original assumption.

(c) Yes. A_n converges to 0 in probability. As n becomes very large, the expected value of A_n approaches 0, and the variance of A_n approaches 0. So, by the Chebychev Inequality, A_n converges to 0 in probability. You could also show this by noting that the A_ns are i.i.d. with finite mean and variance and using the WLLN.

7. (a) Suppose $Y_1, Y_2, \ldots$ converges to a in mean of order p. This means that $\mathbf{E}[|Y_n - a|^p] \to 0$, so to prove convergence in probability we should upper bound $\mathbf{P}(|Y_n - a| \geq \epsilon)$ by a multiple of $\mathbf{E}[|Y_n - a|^p]$. This connection is provided by the Markov inequality.
Let $\epsilon > 0$ and note the bound

$$\mathbf{P}(|Y_n - a| \geq \epsilon) \;=\; \mathbf{P}(|Y_n - a|^p \geq \epsilon^p) \;\leq\; \frac{\mathbf{E}[|Y_n - a|^p]}{\epsilon^p},$$

where the first step is a manipulation that does not change the event under consideration and the second step is the Markov inequality applied to the random variable $|Y_n - a|^p$. Since the inequality above holds for every n,

$$\lim_{n\to\infty} \mathbf{P}(|Y_n - a| \geq \alpha) \;\leq\; \lim_{n\to\infty} \frac{E[|Y_n - a|^p]}{\alpha^p} \;=\; 0.$$

Hence, we have that $\{Y_n\}$ converges in probability to a.

(b) Consider the sequence $\{Y_n\}_{n=1}^{\infty}$ of random variables where

$$Y_n = \begin{cases} 0, & \text{with probability} 1 - \frac{1}{n}; \\ n, & \text{with probability} \frac{1}{n} \end{cases}$$

Note that $\{Y_n\}$ converges in probability to 0, but $E[|Y_n - 0|^1] = E[Y_n|] = 1$ for all n. Hence, $\{Y_n\}$ converges in probability to 0 but not in mean of order 1.

G1†. (a) $\mathbf{E}[\hat{\mu}] = \mathbf{E}[\frac{1}{n}(X_1 + \cdots + X_n)] = \frac{1}{n}(\mathbf{E}[X_1] + \cdots + \mathbf{E}[X_n]) = \frac{1}{n} \cdot n\mathbf{E}[X] = \mu$.
Hence, $\hat{\mu}$ is an an unbiased estimator for the true mean μ.

(b)

$$E[\hat{\sigma}^2] \ = \ \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[(X_i - \mu)^2] = \frac{1}{n} \cdot n\sigma^2 = \sigma^2.$$

Therefore $\hat{\sigma}^2$ (which uses the true mean) is unbiased estimator for σ^2.

(c)

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \hat{\mu})^2 &= \sum_{i=1}^{n}[X_i - \mu - (\hat{\mu} - \mu)]^2 \\
&= \sum_{i=1}^{n}\left[(X_i - \mu)^2 + (\hat{\mu} - \mu)^2 - 2(X_i - \mu)(\hat{\mu} - \mu)\right] \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 + n(\hat{\mu} - \mu)^2 - 2(\hat{\mu} - \mu)\sum_{i=1}^{n}(X_i - \mu) \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 + n(\hat{\mu} - \mu)^2 - 2(\hat{\mu} - \mu)n(\hat{\mu} - \mu) \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\hat{\mu} - \mu)^2
\end{aligned}$$

(d)

$$\begin{aligned}
\mathbf{E}\left[\sum_{i=1}^{n}(X_i - \hat{\mu})^2\right] &= \mathbf{E}\left[\sum_{i=1}^{n}(X_i - \mu)^2\right] - n\mathbf{E}[(\hat{\mu} - \mu)^2] \\
&= n\sigma^2 - n\mathbf{E}\left[\frac{1}{n^2}\left(\sum_{i=1}^{n} X_i - \mu\right)^2\right] \\
&= n\sigma^2 - \frac{1}{n}\mathbf{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - \mu)(X_j - \mu)\right] \\
&= n\sigma^2 - \frac{1}{n}\mathbf{E}\left[\sum_{i=1}^{n}(X_i - \mu)^2\right] \\
&= (n-1)\sigma^2
\end{aligned}$$

where we used the fact that for $i \neq j$, $\mathbf{E}[(X_i - \mu)(X_j - \mu)] = 0$; and for $i = j$, it is is equal to σ^2.

(e) From part (d),

$$\hat{\hat{\sigma}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

is an unbiased estimator for the variance.

(f)

$$\begin{aligned} \text{var}(\hat{\mu}) &= \text{var}(\frac{1}{n}(X_1 + \cdots + X_n)) \\ &= \frac{1}{n^2}(\text{var}(X_1) + \cdots + \text{var}(X_n)) \\ &= \frac{1}{n^2} \cdot n\sigma^2 \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

Thus, $\text{var}(\hat{\mu})$ goes to zero asymptotically. Furthermore, we saw that $\mathbf{E}[\hat{\mu}] = \mu$. Simple application of Chebyshev inequality shows that $\hat{\mu}$ converges in probability to μ (the true mean) as the sample size increases.

(g) Not yet typeset.