Math 466/566 - Homework 2 Solutions

1. Problem 1 from chapter 2 in the book. **Solution:** $E[1_{A_i}] = 2/3$. So $E[1_{A_i}^2] = 2/3$. So $var(1_{A_i}) = 2/3 - (2/3)^2 = 2/9$.

2. Problem 2 from chapter 2 in the book. **Solution:** $\mu_f = 2/3$.

$$\sigma_f = \frac{var(1_{A_i})}{\sqrt{n}} = \frac{1}{15}$$

3. Problem 4 from chapter 2 in the book. **Solution:** $\sigma_f = var(1_{A_i}) = \sqrt{p(1-p)}/\sqrt{n}$. We want this at most 0.01, so $\sqrt{p(1-p)}/\sqrt{n} \leq 0.01$. So $n \geq 10,000p(1-p)$. However, p is unknown, so this is not a very useful answer. The largest $p(1-p)$ can be is 1/4, so if $n \geq 2,500$, this will insure σ_f is at most 0.01.

4. Problem 8 from chapter 2 in the book. **Solution:** The sample space has 36 entries which you can write out. Straightforward but rather tedious counting then shows

$$\begin{aligned} P(A) = P(B) = P(C) &= 1/2 \\ P(A \cap B) = P(A \cap C) = P(B \cap C) &= 1/4 \\ P(A \cap B \cap C) &= 0 \end{aligned}$$

Thus

$$\begin{aligned} P(A \cap B) &= P(A)P(B) \\ P(A \cap C) &= P(A)P(C) \\ P(B \cap C) &= P(B)P(C) \end{aligned}$$

but

$$P(A \cap B \cap C) \neq P(A)P(B)P(C)$$

5. Recall that events A and B are independent if $P(A \cap B) = P(A)\,P(B)$. As we observed in class, if the random variables 1_A and 1_B are independent, then the events A and B are independent. Prove the converse: if the events A and

B are independent, then the random variables 1_A and 1_B are independent. You must show that for any functions $g(x)$ and $h(x)$,

$$E[g(1_A)h(1_B)] = E[g(1_A)]\, E[h(1_B)]$$

Solution: Note that a RV of the form $g(1_A)$ only takes on two values. When ω is in A, $g(1_A) = g(1)$. And when ω is not in A, $g(1_A) = g(0)$. So we have

$$g(1_A) = g(0)\, 1_{A^c} + g(1)\, 1_A = g(0)\,(1 - 1_A) + g(1)\, 1_A = g(0) + (g(1) - g(0))\, 1_A$$

Thus

$$\begin{aligned}
& E[g(1_A)h(1_B)] = E[(g(0) + (g(1) - g(0))\, 1_A)\,(h(0) + (h(1) - h(0))\, 1_B)] \\
= \; & g(0)h(0) + g(0)(h(1) - h(0))E[1_B] \\
+ \; & (g(1) - g(0))h(0)E[1_A] + (g(1) - g(0))(h(1) - h(0))E[1_A 1_B] \\
= \; & g(0)h(0) + g(0)(h(1) - h(0))E[1_B] \\
+ \; & (g(1) - g(0))h(0)E[1_A] + (g(1) - g(0))(h(1) - h(0))E[1_A]E[1_B] \\
= \; & E[g(0) + (g(1) - g(0))\, 1_A]\, E[h(0) + (h(1) - h(0))\, 1_B] \\
= \; & E[g(1_A)]\, E[h(1_B)]
\end{aligned}$$

6. Let X be a random variable with finite mean and variance. Prove that for all constants c,

$$E[(X - c)^2] \geq E[(X - \mu_X)^2]$$

Solution: First note that

$$E[(X - c)^2] = E[X^2] - 2cE[X] + c^2 = E[X^2] - 2c\mu_X + c^2$$

and

$$E[(X - \mu_X)^2] = E[X^2] - 2\mu_X E[X] + \mu_X^2 = E[X^2] - \mu_X^2$$

So the inequality to be proved is equivalent to

$$E[X^2] - 2c\mu_X + c^2 \geq E[X^2] - \mu_X^2$$

which is equivalent to

$$\mu_X^2 - 2c\mu_X + c^2 \geq 0$$

which is true for all c since $\mu_X^2 - 2c\mu_X + c^2 = (\mu_X - c)^2$.
Solution for 566:
566 students had to show the median minimizes $E[|X-c|]$. For any c,

$$\begin{aligned} E[|X-c|] &= \int_c^\infty (x-c)f(x)dx + \int_{-\infty}^c (c-x)f(x)dx \\ &= \int_c^\infty xf(x)dx - \int_{-\infty}^c xf(x)dx + c[P(X \le c) - P(X \ge c)] \end{aligned}$$

In particular, using this with $c = m$ for which $P(X \le m) = P(X \ge m)$, we have

$$E[|X-m|] = \int_m^\infty xf(x)dx - \int_{-\infty}^m xf(x)dx$$

Thus

$$\begin{aligned} E[|X-c|] - E[|X-m|] &= 2\int_c^m xf(x)dx + c[P(X \le c) - P(X \ge c)] \\ &= 2\int_c^m xf(x)dx + c[2P(X \le c) - 1] \end{aligned}$$

Now differentiate this with respect to c and we get

$$-2cf(c) + [2P(X \le c) - 1] + c2f(c) = 2P(X \le c) - 1 \tag{1}$$

This is only zero when $P(X \le c) = 1/2$, i.e., when c equals the median. As a function of c the quantity $E[|X-c|] - E[|X-m|]$ goes to ∞ when c goes to ∞ or $-\infty$, so the critical point must be a minimum.

7. Let $A_1, A_2, \cdots, A_n$ be independent events with the same probability p. We studied the problem of estimating the population proportion p by using the sample proportion. We saw that $f_n = \bar{X}_n$, i.e., the sample proportion is equal to the sample mean of the random variables $1_{A_1}, \cdots, 1_{A_n}$. This is an estimator for the population proportion, p. The variance of the random variable 1_{A_i} is $p(1-p)$, but p is unknown. Since $\bar{X}_n$ is hopefully close to p, a possible estimator for the variance $p(1-p)$ is $\bar{X}_n(1-\bar{X}_n)$. Show that the mean of this estimator is

$$E[\bar{X}_n(1-\bar{X}_n)] = p(1-p)\frac{n-1}{n}$$

Solution:

$$E[\bar{X_n}(1-\bar{X_n})] = E[\bar{X_n}^2] - E[\bar{X_n}]$$

We know $E[\bar{X_n}] = p$. We also know that $var(\bar{X_n}) = p(1-p)/n$. Since $var(\bar{X_n}) = E[\bar{X_n}^2] - E[\bar{X_n}]^2$, this implies $E[\bar{X_n}^2] = p(1-p)/n + p^2$. So

$$E[\bar{X_n}(1-\bar{X_n})] = p(1-p)/n + p^2 - p = \frac{n-1}{n}p(1-p)$$