Chapter 7.4: Expected Value and Variance
Monday, August 3

Summary

- Indictor Variables: \( I_E(s) = \begin{cases} 1 & s \in E \\ 0 & s \notin E \end{cases} \)
- Variance: \( \text{Var}(X) = E((X - E(X))^2) = E(X^2) - E(X)^2 \)
- If \( X \) and \( Y \) are independent then \( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \)
- If \( X_1, \ldots, X_n \) are pairwise independent then \( \text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i) \).
- \( \text{Var}(aX + b) = a^2 \cdot \text{Var}(X) \).
- Markov’s Inequality: If \( X \) is non-negative then \( p(X \geq a) \leq E(X)/a \).
- Chebyshev’s Inequality: \( p(|X - E(X)| \geq r) \leq \text{Var}(X)/r^2 \)
- Covariance: \( \text{Cov}(X,Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y) \)
- \( \text{Cov}(X,Y) = \text{Cov}(Y,X), \text{Cov}(X,Y + Z) = \text{Cov}(X,Y) + \text{Cov}(X,Z) \)
- \( E \) and \( F \) are positively correlated if \( p(E \cap F) > p(E)p(F) \).

Variance

1. (★) If \( X \) is the sum of a rolled pair of dice, what is \( \text{Var}(X) \)?
   
   Let \( D_1 \) and \( D_2 \) be the separate dice rolled. Then \( \text{Var}(X) = \text{Var}(D_1 + D_2) = \text{Var}(D_1) + \text{Var}(D_2) \) since the two rolls are independent. So we just have to compute the variance of a single roll of the die:
   
   \[
   \begin{align*}
   (a) \text{ Method 1: } \text{Var}(D_1) &= E(D_1^2) - (E(D_1))^2 = \frac{1}{6}((-2.5)^2 + (-1.5)^2 + (.5)^2 + .5^2 + 1.5^2 + 2.5^2) = 35/12. \\
   (b) \text{ Method 2: } \text{Var}(D_1) &= E(D_1^2) - (E(D_1))^2 = \frac{1}{6}1 + 4 + 9 + 16 + 25 + 36 - 12.25 = 35/12.
   \end{align*}
   \]

   So the variance of \( X \) is \( 2 \cdot (35/12) = 35/6 \).

2. (★) If a coin with a 25% chance of landing on heads and \( X \) is the number of heads that result from 50 flips of the coin, what is \( \text{Var}(X) \)?

   The variance of a single flip is \( p(1-p) = (1/4)(3/4) = 3/16 \). Then since the 50 flips are independent, the variance is \( 50 \cdot (3/16) = 75/8 \).

3. If \( E \) is some event, what is \( \text{Var}(I_E) \)?

   \[ \text{Var}(I_E) = E(I_E^2) - E(I_E)^2 = E(I_E) - E(I_E)^2 = p(E) - p(E)^2 = p(E)(1-p(E)) \]

4. Prove by induction: If \( X_1, \ldots, X_n \) are \( \text{mutually} \) independent random variables, prove by induction that \( \text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i) \).

   Base case: If \( X \) and \( Y \) are independent then the identity \( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \) has already been established.
Inductive step: Suppose that \( \text{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{Var}(X_i) \). Then
\[
\text{Var}(\sum_{i=1}^{n+1} X_i) = \text{Var}(\sum_{i=1}^{n} X_i + X_{n+1})
\]
\[
= \text{Var}(\sum_{i=1}^{n} X_i) + \text{Var}(X_{n+1})
\]
\[
= \sum_{i=1}^{n} \text{Var}(X_i) + \text{Var}(X_{n+1})
\]
\[
= \sum_{i=1}^{n+1} \text{Var}(X_i)
\]
Thus the formula holds for all \( n \geq 2 \) by induction. The step from the first line to the second comes from the fact that the variables are mutually independent (and so \( \sum_{i=1}^{n} X_i \) and \( X_{n+1} \) are independent), and the next step comes from the inductive hypothesis.

5. Why does induction not work if the variables are only pairwise independent?

If \( X, Y, \) and \( Z \) are only pairwise independent it is not necessarily true that \( X + Y \) and \( Z \) are independent, and so we can’t say automatically that \( \text{Var}(X + Y + Z) = \text{Var}(X + Y) + \text{Var}(Z) \).

**Chebyshev’s Inequality**

1. (\( \star \)) If a fair coin is flipped 100 times, use Chebyshev’s inequality to bound the probability of getting at least 55 or at most 45 heads.

   The variance of a single flip is 1/4, so the variance of 100 flips is 25. Therefore \( p(|X - 50| \geq 5) \leq 25/5^2 = 1 \), which is not a very helpful bound because the probability of any event is at most 1.

2. Bound the probability of getting at least 65 or at most 35 heads.

   \( p(|X - 50| \geq 15) \leq 25/15^2 = 1/9 \), so there is at least an 8/9 chance that 100 flips will get between 36 and 64 heads.

3. Prove: if you flip \( N \) fair coins and \( X_N \) is the number of heads, then \( \lim_{n \to \infty} p(|X - N/2| \geq \epsilon N) = 0 \) for any \( \epsilon > 0 \).

   For any \( N \) and \( \epsilon \), \( p(|X_N - N/2| \geq \epsilon N) \leq \frac{N/4}{(N\epsilon)^2} = \frac{1/\epsilon^2}{4N}. \) For any fixed \( \epsilon \), this quantity goes to 0 as \( N \to \infty \).

4. Prove: If \( X_1, X_2, \ldots \) are independent identically distributed (i.i.d.) random variables with finite variance and expected value \( \mu \), and \( \mu_n = \frac{1}{n} \sum_{i=1}^{n} X_i \) (the average of the first \( n \) variables), then \( \lim_{n \to \infty} p(|\mu_n - \mu| > \epsilon) = 0 \) for any \( \epsilon > 0 \).

   Very similar to the previous proof: \( p(|\mu_n - \mu| > \epsilon) = p(|N\mu_n - N\mu| > N\epsilon) \leq N \cdot \text{Var}(X_1)/(N\epsilon)^2 = \frac{\text{Var}(X_1)/\epsilon^2}{N} \), which approaches 0 for any fixed \( \epsilon \) as \( N \to \infty \).
Covariance

1. Show that if $k$ is any constant (or more precisely, a random variable that always takes on the value $k$) then \( \text{Cov}(k, X) = 0 \).

\[
\text{Cov}(k, X) = E(kX) - E(k)E(X) = kE(X) - kE(X) = 0.
\]

2. (★) Show that \( \text{Cov}(aX + b, Y) = a \cdot \text{Cov}(X, Y) \).

\[
\text{Cov}(aX + b, Y) = E(aXY + bY) - E(aX + b)E(Y) = aE(XY) + bE(Y) - aE(X)E(Y) - bE(Y) = aE(XY) - aE(X)E(Y) = a \cdot \text{Cov}(X, Y).
\]

3. If \( p(E) = .6 \) and \( p(F) = .8 \), what can we say about the correlation of \( E \) and \( F \)?

Nothing.

4. (★) Flip 100 coins and let \( X \) be the number of times \( HT \) appears. Find \( \text{Var}(X) \).

Let \( X_i \) be the event that the string \( HT \) appears in the \( i \)-th and \( (i+1) \)-th spots. Then we write \( X = X_1 + X_2 + \cdots + X_{99} \), and expand \( \text{Var}(X) \) in terms of covariance:

\[
\text{Var}(X) = \text{Cov}(\sum_{i=1}^{99} X_i, \sum_{i=1}^{99} X_i)
\]

\[
= \sum_{i=1}^{99} \text{Var}(X_i) + 2 \sum_{i<j} \text{Cov}(X_i, X_j)
\]

If \( j \geq i + 2 \) or \( i \geq j + 2 \), the random variables \( X_i \) and \( X_j \) are independent because the sequences of flips do not overlap. Thus the covariance is non-zero only if \( i - 1 \leq j \leq i + 1 \). Therefore

\[
\text{Var}(X) = \sum_{i=1}^{99} \text{Var}(X_i) + 2 \sum_{i=1}^{98} \text{Cov}(X_i, X_{i+1})
\]

For the first sum, the variances are identical and equal to \((1/4)(3/4) = 3/16\). Each term in the second sum is equal to \( p(HT - \cap - HT) - p(HT -)p(-HT) = 0 - (1/4)(1/4) = -1/16 \) (Note that the probability of \( X_1 = 1 \) and \( X_2 = 1 \) happening simultaneously is zero). The variance is therefore

\[
99 \cdot (3/16) - 2 \cdot 98 \cdot (-1/16) = 101/16 = 6.3125.
\]

The expected number of \( HT \) sequences is \( 99/4 = 24.75 \).

5. Flip 100 coins. Let \( X \) be the number of times \( HHH \) appears and let \( Y \) be the number of times \( HHT \) appears. Find \( \text{Cov}(X, Y) \).

Write \( X = X_1 + X_2 + \cdots + X_{98} \) and \( Y = Y_1 + \cdots + Y_{98} \) as a sum of indicator variables. Then using the fact that \( X_i \) and \( Y_j \) are independent if \( |i - j| \geq 3 \) (the flips do not overlap), we get

\[
\text{Cov}(X, Y) = \text{Cov}(\sum_i X_i, \sum_j Y_j)
\]

\[
= \sum_{i,j} \text{Cov}(X_i, Y_j)
\]

\[
= \sum_{i=1}^{96} \text{Cov}(X_i, Y_{i+2}) + \sum_{i=1}^{97} \text{Cov}(X_i, Y_{i+1}) + \sum_{i=1}^{98} \text{Cov}(X_i, Y_i) + \sum_{i=2}^{98} \text{Cov}(X_i, Y_{i-1}) + \sum_{i=3}^{98} \text{Cov}(X_i, Y_{i-2})
\]

\[
= 96 \cdot \text{Cov}(X_1, Y_3) + 97 \cdot \text{Cov}(X_1, Y_2) + 98 \cdot \text{Cov}(X_1, Y_1) + 97 \cdot \text{Cov}(X_2, Y_1) + 96 \cdot \text{Cov}(X_3, Y_1)
\]

Since these are indicator variables, we know that \( \text{Cov}(X_i, Y_j) = E(X_iY_j) - E(X_i)E(Y_j) = P(X_i \cap Y_j) - P(X_i)P(Y_j) = P(X_i \cap Y_j) - 1/64 \) (since \( X_i \) and \( Y_j \) are both specific three-coin sequences).
The probability of $Y_1$ occurring at the same time as $X_1$, $X_2$, or $X_3$ is zero. The probability of $Y_2$ occurring at the same time as $X_1$ is $1/16$, and the probability of $Y_3$ occurring at the same time as $X_1$ is $1/32$. The covariance is therefore

$$Cov(X, Y) = 96 \cdot (1/32 - 1/64) + 97 \cdot (1/16 - 1/64) + 98 \cdot (-1/64) + 97 \cdot (-1/64) + 96 \cdot (-1/64) = 1.5$$

6. Prove that if $P(E|F) > P(E)$ then $P(E|\overline{F}) < P(E)$ (in other words, if $E$ and $F$ are positively correlated then $E$ and $\overline{F}$ are negatively correlated).

$$P(E) = P(E|F)P(F) + P(E|\overline{F})P(\overline{F})$$
$$> P(E)P(F) + P(E|\overline{F})P(\overline{F})$$
$$P(E)(1 - P(F)) > P(E|\overline{F})P(\overline{F})$$
$$P(E)P(\overline{F}) > P(E|\overline{F})P(\overline{F})$$
$$P(E) > P(E|\overline{F})$$