Notes for Recitation 12 (12/4/98)

1 Conditional Expectation

We can define *conditional expectation*, or the expectation of a random variable, conditioned on a particular event as follows:

Definition 1.1. The conditional expectation of a random variable R given an event A is defined as

$$E[R \mid A] = \sum_{r \in Range(R)} r \cdot Pr(R = r \mid A).$$

Example: Let R be the outcome of a die. What is the expected value of R given that $R \leq 4$?

$$E[R \mid R \leq 4] = \sum_{r=1}^{4} r Pr(R = r \mid R \leq 4)$$

But $Pr(R = r \mid R \leq 4) = \frac{1}{4}$. So,

$$E[R \mid R \leq 4] = \frac{1}{4} \sum_{r=1}^{4} r = 2.5$$

One use of conditional expectation is to compute an expectation.
The following theorem breaks down the expectation into conditional expectations.

Theorem 1.2. Let $A_1, A_2, \ldots, A_n$ be a set of disjoint events whose union is the the sample space. Then for any random variable R,

$$E[R] = \sum_{i=1}^{n} E[R \mid A_i] Pr(A_i)$$

Notice that this is analogous to the theorem of total probability:

Theorem 1.3. Let $A_1, A_2, \ldots, A_n$ be a set of disjoint events whose union is the the sample space. Then for any event B,

$$Pr(B) = \sum_{i=1}^{n} Pr(B \cap A_i).$$

Proof of Theorem 1.

$$\begin{aligned}
E[R] &= \sum_r r \cdot Pr(R = r) \\
&= \sum_r r \cdot \left(\sum_{i=1}^{n} Pr(R = r \cap A_i) \right) \\
&= \sum_r r \cdot \left(\sum_{i=1}^{n} Pr(R = r \mid A_i) Pr(A_i) \right) \\
&= \sum_{i=1}^{n} \left(\sum_r r \cdot Pr(R = r \mid A_i) \right) Pr(A_i) \\
&= \sum_i E[R \mid A_i] Pr(A_i)
\end{aligned}$$

The first step uses the theorem of total probability with B being the event $R = r$. The second step is by the definition of conditional probability. □

Example: (Assembly plants)

Suppose there are two computer assembly plants F and S. Plant F is fast but sloppy: It makes 70% of the computers, but a computer from F has a 2% probability of breaking down on any given day of use.

Plant S is slow and careful: It makes 30% of the computers, but the breakdown probability is only 1%.

These computers get randomly mixed, so I buy one at random from the store. What is the expected number of days my computer works?

Let D be the random variable giving the number of days my computer works. Let I_S be the event that I bought the computer from S, and let I_F be that I bought it from F. We solve this problem using conditional expectations. We calculate separately the conditional expectation of the number of working days conditioned on I_S and on I_F, then combine using the conditional expectation theorem.

The expected number of working days for a computer from S with probability of failure $p = 0.01$ is just

$$E[D \mid I_S] = \frac{1}{p} - 1 = 99$$

Note that mean time to failure is $\frac{1}{p}$, hence the number of working days is $\frac{1}{p} - 1$. Similarly, the expected number of working days for a computer from F is

$$E[D \mid I_F] = \frac{1}{.02} - 1 = 49$$

.

Now to calculate the expected number of working days for the computer when we don't know which factory it came from, we use the conditional expectation theorem:

$$E[D] = E[D|I_S]Pr(I_S) + E[D|I_F]Pr(I_F) = 99(.30) + 49(.70) = 64$$

2 Covariance

We saw in class that the variance of a random variable is a measure of how it deviates from the mean. *Covariance* is another statistical measure. It quantifies the similarity of the distributions of two random variables.

Definition 2.1. The covariance of random variables R_1 and R_2 is

$$\mathrm{Cov}\left[R_1, R_2\right] ::= \mathrm{E}\left[R_1 R_2\right] - \mathrm{E}\left[R_1\right]\mathrm{E}\left[R_2\right].$$

If the covariance of R_1 and R_2 is positive, we will say that they are "positively correlated", and if their covariance is negative we say that they are "negatively correlated". Intuitively, positively correlated random variables tend to have close values, while negatively correlated variables tend to have values far apart. Independent random variables have zero correlation.

Examples: Let R_1 and R_2 be uniform in the range $[1, 6]$.

1. $R_1 = R_2$. In this case,

   $$\mathrm{Cov}\left[R_1, R_2\right] = \mathrm{E}\left[R_1^2\right] - \mathrm{E}^2\left[R_1\right] = \mathrm{Var}\left[R_1\right] = 2\frac{11}{12}.$$

   The variables are positively correlated. This is not a surprise — they are always equal.

2. $R_2 = 7 - R_1$. Now,

   $$\begin{aligned} \mathrm{Cov}\left[R_1, R_2\right] &= \mathrm{E}\left[R_1(7 - R_1)\right] - \mathrm{E}\left[R_1\right]\mathrm{E}\left[7 - R_1\right] \\ &= 7\mathrm{E}\left[R_1\right] - \mathrm{E}\left[R_1^2\right] - 7\mathrm{E}\left[R_1\right] + \mathrm{E}^2\left[R_1\right] \\ &= -\mathrm{E}\left[R_1^2\right] + \mathrm{E}^2\left[R_1\right] \\ &= -\mathrm{Var}\left[R_1\right] = -2\frac{11}{12} \end{aligned}$$

   The variables are negatively correlated, i.e., when one is high the other is low. This is reflected by the negative covariance.

3. R_1 are R_2 are independent (e.g., consecutive coin tosses). Then

$$\mathrm{Cov}[R_1, R_2] = \mathrm{E}[R_1 R_2] - \mathrm{E}[R_1]\,\mathrm{E}[R_2] = \mathrm{E}[R_1]\,\mathrm{E}[R_2] - \mathrm{E}[R_1]\,\mathrm{E}[R_2] = 0.$$

   The correlation of independent random variables is zero.

So if the variables are independent then, indeed their covariance is zero. What about the converse? If the covariance is zero, are the variables independent? The answer is no. For example, let R_1, R_2 be in the range $[1,3]$ such that $\Pr[R_1 = i \& R_2 = j] = 1/8$ for all (i,j) except $(2,2)$ and $\Pr[R_1 = R_2 = 2] = 0$. We can check that this gives valid probability distributions for R_1 and R_2:

$$\Pr[R_1 = 1] = \frac{3}{8},\ \Pr[R_1 = 2] = \frac{2}{8},\ \Pr[R_1 = 3] = \frac{3}{8},$$

which add up to 1.

The distribution of R_1 is the same as the distribution of R_2 but since $\Pr[R_1 = 2 \mid R_2 = 2] = 0 \neq \Pr[R_1 = 2]$, R_1 and R_2 are not independent. However,

$$\begin{aligned}
\mathrm{Cov}[R_1, R_2] &= \mathrm{E}[R_1 R_2] - \mathrm{E}[R_1]\,\mathrm{E}[R_2] \\
&= \sum_{i,j\in[1,3]} ij \Pr[R_1 = i \& R_2 = j] - \Big(\sum_{i\in[1,3]} i \Pr[R_1 = i]\Big)^2 \\
&= \frac{1}{8}(1+2+3+2+6+3+6+9) - \left(\frac{3}{8}(1+3) + \frac{2}{8}2\right)^2 \\
&= 4 - 4 = 0
\end{aligned}$$

Theorem 2.2. For random variables R_1 and R_2,

$$\mathrm{Var}[R_1 + R_2] = \mathrm{Var}[R_1] + \mathrm{Var}[R_2] + 2\mathrm{Cov}[R_1, R_2].$$

Proof.

$$\begin{aligned}
\mathrm{Var}[R_1 + R_2] &= \mathrm{E}\left[(R_1 + R_2)^2\right] - \mathrm{E}^2[R_1 + R_2] \\
&= \mathrm{E}\left[R_1^2 + 2R_1R_2 + R_2^2\right] - (\mathrm{E}[R_1] + \mathrm{E}[R_2])^2 \\
&= \mathrm{E}\left[R_1^2\right] + 2\mathrm{E}[R_1R_2] + \mathrm{E}\left[R_2^2\right] - \mathrm{E}^2[R_1] - 2\mathrm{E}[R_1]\,\mathrm{E}[R_2] - \mathrm{E}^2[R_2] \\
&= \left(\mathrm{E}\left[R_1^2\right] - \mathrm{E}^2[R_1]\right) + \left(\mathrm{E}\left[R_2^2\right] - \mathrm{E}^2[R_2]\right) + 2(\mathrm{E}[R_1R_2] - \mathrm{E}[R_1]\,\mathrm{E}[R_2]) \\
&= \mathrm{Var}[R_1] + \mathrm{Var}[R_2] + 2\mathrm{Cov}[R_1, R_2]
\end{aligned}$$

□

3 Variance of Sums

As a corollary to the previous theorem we have:

Corollary 3.1. If R_1 and R_2 are independent, then

$$\mathrm{Var}\left[R_1 + R_2\right] = \mathrm{Var}\left[R_1\right] + \mathrm{Var}\left[R_2\right]$$

Proof. If R_1, R_2 are independent, then $\mathrm{Cov}\left[R_1, R_2\right] = 0$. □

A similar result holds for the sum of many independent random variables — but only needs that the variables are *pairwise* independent.

Theorem 3.2. If $R_1, \ldots, R_n$ are pairwise independent random variables, then

$$\mathrm{Var}\left[R_1 + \cdots + R_n\right] = \mathrm{Var}\left[R_1\right] + \cdots + \mathrm{Var}\left[R_n\right].$$

The proof is simple and is left to the reader. (Calculations will be easier if one defines new random variables $D_i ::= R_i - \mathrm{E}\left[R_i\right]$. Then $\mathrm{Var}\left[\sum_i R_i\right] = \mathrm{E}\left[\left(\sum_i D_i\right)^2\right]$, $\mathrm{E}\left[D_i\right] = 0$, and $\mathrm{Var}\left[R_i\right] = \mathrm{E}\left[D_i^2\right]$, which will simplify the expressions.)

Using this result we can now derive the variance of a binomial distribution. Let R have an n, p binomial distribution, i.e.,

$$\Pr\left[R = k\right] = \binom{n}{k} p^k (1-p)^{n-k}.$$

We can think of R as the sum of n Bernoulli independent random variables. That is, the number of heads in n tosses with the probability of heads equal to p:

$$R = R_1 + R_2 + \ldots + R_n, \ \Pr\left[R_i = 1\right] = p, \ \text{and} \ \Pr\left[R_i = 0\right] = 1 - p.$$

Now

$$\mathrm{Var}\left[R_i\right] = \mathrm{E}\left[R_i^2\right] - \mathrm{E}^2\left[R_i\right] = p - p^2.$$

Hence from the theorem,

$$\mathrm{Var}\left[R\right] = \sum_{i=1}^{n} \mathrm{Var}\left[R_i\right] = n\mathrm{Var}\left[R_i\right] = p(1-p)n.$$

4 Sum of Pairwise Independent Random Variables

Let $V_1, \ldots, V_n$ be pairwise independent random variables such that $\Pr[V_i = 0] = 1/2$ and $\Pr[V_i = 1] = 1/2$ (i.e. each V_i is an uniformly distributed Bernoulli variable.) Then $\text{Var}[\sum_i V_i] = \sum_i \text{Var}[x_i] = n/4$, and Chebyshev inequality implies that

$$\Pr\left[\left|\sum_i V_i - n/2\right| \geq \epsilon n\right] \leq \frac{1}{4\epsilon^2 n}.$$

In particular, for $\epsilon = 1/2$,

$$\Pr\left[\sum_i V_i = 0 \text{ or } \sum_i V_i = n\right] \leq \frac{1}{n}.$$

If the random variables V_i are mutually independent, then the previous bound can be strengthened to $2 \cdot 2^{-n}$. We will see that this is not possible in general, i.e., under the sole condition of pairwise independence, the bound implied by Chebyshev's inequality is essentially the best possible.

Consider independent uniformly distributed Bernoulli variables $X_1, \ldots, X_k$. Let $n ::= 2^k - 1$; we shall define n new random variables Y_S, one for any non-empty subset $S \subseteq \{1, \ldots, k\}$, as follows:

$$Y_S = \left(\begin{array}{ll} 1 & \text{if } \sum_{i \in S} X_i \text{ is odd,} \\ 0 & \text{otherwise.} \end{array}\right.$$

It is easy to verify that for every non-empty S, the random variable Y_S is a uniformly distributed Bernoulli random variable (Y_S can be thought of as the indicator variable of the event "make k consecutively numbered coin tosses and get an odd number of heads among the coin tosses whose numbers are in S." Such an event happens with probability $1/2$ as long as S is not empty.)

We now proceed to show that for every two different non-empty set $S \neq T \subseteq \{1, \ldots, k\}$, the random variables Y_S and Y_T are independent. Fix two boolean values $b_1, b_2 \in \{0, 1\}$, we want to show that

$$Pr(Y_S = b_1 | Y_T = b_2) = Pr(Y_S = b_1)$$

5 Team Problems

Problem 1. The Petersburg Casino offers the following game: the gambler bets a fixed wager, and then the dealer flips a fair coin (dealers do not flip coins in US casinos, but they do in Petersburg) until it comes up heads. The gambler receives \$1 if the coin shows heads the first time, \$2 if it shows the first head at second toss, and in general $\$2^{k-1}$ if the dealer tosses the coin k times to get the first head. The fixed wager is \$100,000.

(a) What is the expected amount of money that the Petersburg Casino is going to pay to a gambler that plays this game? What about the variance? Is it reasonable for the gambler to risk \$100,000 dollars in a game where there is a 50% chance of getting only *one dollar* back (this is not really a mathematical question: just think about that.)

(b) What is the probability that the gambler does not lose money in this game, i.e. what is the probability that he gets back from the Casino more than \$100,000?

Problem 2. A Harvard student, lost in the underground corridors of MIT, finds himself in a room containing three doors. Door 1 leads him to freedom (out of MIT) after an hour of travel; door 2 returns him to the same room after four hours travel, and door 3 returns him to the room after seven hours travel. Each time he finds himself in the room, the Harvard student picks a door with equal probability. Let T denote the time until he finds the exit.

(a) Define a sequence of independent identically distributed random variables R_i and a stopping rule Q such that

$$T = \sum_{i=1}^{Q} R_i$$

(b) Argue that R_k is independent of the event that $Q \geq k$. Is R_k independent of Q?

(c) Use Wald's Theorem to find $\mathrm{E}[T]$.

(d) Now suppose that an MIT student is also lost and finds herself in a similar situation to the Harvard student. The MIT student never chooses the same door twice. What is $\mathrm{E}[T]$ for the MIT student?

6 Solutions

Problem 1 Solution

Let V be the random variable corresponding to the amount of money that the gambler is paid by the dealer. The distribution of V is as follows: for any $n \geq 1$,

$$\Pr\left[V = 2^{n-1}\right] = 2^{-n},$$

where the probability refers to the event that the dealer tosses $n-1$ tails followed by a head. The average of V is

$$\mathrm{E}\left[V\right] = \sum_{n=1}^{\infty}[2^{-n}2^{n-1}] = \sum_{n=1}^{\infty} 1/2 = \infty.$$

The variance is not defined for a random variable having infinite expectation. It does not sound too bad to risk \$100,000 on a game where we expect to win an infinite amount of money. However the probability of recovering the money of the wager is only

$$\begin{aligned}
\Pr\left[X \geq 100,000\right] &= \Pr\left[X \geq 2^{17}\right] \\
&= \Pr\left[\text{ there are 17 consecutive tails }\right] \\
&= \frac{1}{131,072}
\end{aligned}$$

Problem 2 Solution

We view this an experiment where at each step the student makes a choice between three equally likely outcomes. We define R_i as the number of hours to return to the same room, or to escape, starting at the ith step. The experiment does not end. That is, as soon as the student escapes, he is doomed, like Sisyphus, to be stuck back in the room and start trying to escape again. This ensures that there is an ith step for all $i \geq 0$, and that the R_i are identically distributed (they are also independent, though this fact is not needed in applying Wald's Theorem). In particular, $\Pr\left[R_i = 1\right] = \Pr\left[R_i = 4\right] = \Pr\left[R_i = 7\right] = 1/3$, and $\mathrm{E}\left[R_i\right] = (1+4+7)/3 = 4$.

Define Q to be the number of steps from the first step until the escape door is selected *for the first time*. Now Q is geometrically distributed with probability $1/3$ and $\mathrm{E}\left[Q\right] = 3$.

Assuming Wald's Theorem is applicable, we conclude

$$\mathrm{E}\left[T\right] = \mathrm{E}\left[\sum_{i=1}^{Q} R_i\right] = 3 \times 4 = 12.$$

To justify the use of Wald's Theorem, we establish that R_i is independent of the event that $Q \geq i$.

It's worth noting that R_i is *not* independent of Q. For example, if $Q = i$, then it must be that the student found the escape door at the ith step, and so $\Pr[R_i = 1 | Q = i] = 1 \neq 1/3 = \Pr[R_i = 1]$.

However, suppose we are given that the student has never managed to escape before the ith step, that is, that $Q \geq i$. Then we can observe that the experiment starting at the ith step has exactly the same structure as the experiment at the first step—where we were given, implicitly, that the student had not escaped before the first step. So we can conclude that the distribution of R_i given $Q \geq i$ is the same as that of R_1 given that $Q \geq 1$, that is, the same as R_1, which we already observed is the same as R_i. In other words, R_i is independent of the event that $Q \geq i$, as required. (This informal argument can be justified rigorously as in Handout 54 on "Probabilities in a Dice Game," by explicitly describing the sample space for the problem along with a suitable bijection between the whole sample space and the event $Q \geq i$.)

Regarding the MIT student, the problem can be solved by complete enumeration; call A the door that leads out, B the door that leads back in four hours, and C the door that leads back in seven hours. The possible events are

- A with probability $1/3$, in which case $T = 1$
- BA with probability $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$, in which case $T = 5$;
- BCA with probability $\frac{1}{2} \cdot \frac{1}{3} \cdot 1 = \frac{1}{6}$, in which case $T = 12$;
- CA with probability $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$, in which case $T = 8$;
- CBA with probability $\frac{1}{2} \cdot \frac{1}{3} \cdot 1 = \frac{1}{6}$, in which case $T = 12$.

The average of T this time is $\mathrm{E}[T] = 39/6 = 6.5$.