21-228 Week 9 notes

November 2, 2001

1 More Expectation

1.1 Conditional Expectation

Definition 1.1. *Suppose $A \subseteq \Omega$ and Z is a random variable on Ω Then*

$$E(Z|A) = \sum_{\omega \in A} Z(\omega)P(\omega|A) = \sum_{k \in \mathbb{R}} kP(Z = k|A)$$

The law of total probability that we discussed earler today has an analog with expectations:

Theorem 1.2. *Let $B_1, B_2, \ldots, B_n$ be a collection of events partitionaing Ω. Then $E(Z) = \sum_{i=1}^{n} E(Z|B_i)\Pr(B_i)$*

Proof.

$$\begin{aligned}
\sum_{i=1}^{n} E(Z|B_i)P(B_i) &= \sum_{i=1}^{n} \sum_{\omega \in B_i} Z(\omega)\frac{P(\omega)}{P(B_i)}P(B_i) \\
&= \sum_{i=1}^{n} \sum_{\omega \in B_i} Z(\omega)P(\omega) \\
&= \sum_{\omega \in \Omega} Z(\omega)P(\omega) \\
&= E(Z)
\end{aligned}$$

□

1.1.1 Two Dice

Let A be the event $x_1 \geq 3$. What is $E(Z|A)$? There are 24 total outcomes where $x_1 \geq 3$:

We get:

$$\begin{aligned} E(Z) &= 4\left(\frac{1}{24}\right) + 5\left(\frac{2}{24}\right) + 6\left(\frac{3}{24}\right) + 7\left(\frac{4}{24}\right) + \cdots + 12\left(\frac{1}{24}\right) \\ &= \left(\frac{182}{24}\right) = 7\frac{7}{12} \end{aligned}$$

1.1.2 Hashing

Let $U = \{0, 1, \ldots, N-1\}$ and $H = \{0, 1, \ldots, n-1\}$. Assume that $n|N$ and that N is much larger than n. Define $f : U \to H$ by $f(u) = u \bmod n$ (that is u maps to whatever its remainder is when divided by n.

Think of H as a hash table and U the universe of objects from which a subset is to be stored. The idea of hashing, in general, is to "classify" the objects so that when we look for them later, it is much easier to find them. The elements of H are called *cells*.

Suppose we with to place objects $u_1, \ldots, u_m$ from U in the hash table, where $m = \alpha n$. Assume that this subset of m elements is randomly chosen from all subsets of m elements.

We store elements in the hash table by placing element u in cell $f(u)$. Elements of a given cell are stored in a linked list – so if an element u is stored in cell h, and there are k elements total in cell h, we may need up to k comparisons to get to element u.

Question: Suppose that we uniformly randomly pick an element u from the has table.

(1) Let T_1 be the number of comparisons needed to find U in the table. What is $E(X)$?

Answer:

Let $M = N/n$, the number of u's that map to a given cell. Let X_k denote the number of u_i for which $f(u_i) = k$. Then by the (expectation) law of total probability we have:

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{n-1} E(T_1|f(u)=k)\Pr(f(u)=k) \\
&= \frac{1}{n}\sum_{k=0}^{n-1} E(T_1|f(u)=k) \\
&= \frac{1}{n}\sum_{k=0}^{n-1} E\left(\frac{X_k}{M}\frac{1+X_k}{2} + \left(1-\frac{X_k}{M}\right)X_k\right) \\
&= \frac{1}{n}\sum_{k=0}^{n-1} E\left(X_K\left(\frac{1+X_k}{2M}+1-\frac{X_k}{M}\right)\right) \\
&= \frac{1}{n}\sum_{k=0}^{n-1} E\left(X_k\left(1-\frac{X_k-1}{2M}\right)\right)
\end{aligned}
$$

If $X_k = 0$, we know no comparisons are needed, and if X_k is at least 1, then we know that $1 - \frac{X_k-1}{2M}$ is at most 1. Therefore, the quantity above is at most:

$$
\begin{aligned}
\frac{1}{n}\sum_{k=0}^{n-1} E(X_k) &= E\left(\sum_{k=0}^{n-1} X_k\right) \\
&= \frac{1}{n}m \\
&= \frac{1}{n}\alpha n \\
&= \alpha
\end{aligned}
$$

So $E(T_1) \leq \alpha$.

1.1.3 Random Walk

Let us return to the example of a particle on a line, where the particle makes n moves total.

Question: What is the expected number of times that the particle visits the origin (counting the fact that it starts there as a visit)?

Solution: Let X_i be the random variable's position after i moves. let $Y_i = 1$ if $X_i = 0$ and $Y_i = 0$ otherwise. Let Y be the number of times that the particle visits the origin. Then $Y = Y_1 + \cdots + Y_n$. If i is odd, we know that Y_i must be 0. If i is even, then we know that $E(Y_i) = \binom{i}{i/2}2^{-i}$.

So $E(Y) = \sum_{0 \leq m \leq n} \binom{m}{m/2}2^{-m}$.

Using Stirling's formula, we can show that each term is approximately $\frac{2}{\pi m}$. We can approximate a sum by taking an integral, but here we have to divide by 2 since we only want to sum over the even numbers. So we get $\sqrt{2n/\pi}$.

2 Independent Random Variables

Definition 2.1 (Independent Random Variables). *Random variables X, Y on Ω are called independent if for all $\alpha, \beta \in \mathbb{R}$, the events $X = \alpha$ and $Y = \beta$ are independent.*

As you'll see in HW6, if we have n random bits, and two variables X and Y depend only on the bits in disjoint sets A, B, respectively, then X, Y are independent.

Here is a nifty theorem:

Theorem 2.2. *If X, Y are $\underline{\text{independent}}$ random variables over Ω, then $E(X \cdot Y) = E(X)E(Y)$.*

Proof.

$$\begin{aligned}
E(XY) &= \sum_\alpha \sum_\beta \alpha\beta \Pr(X = \alpha, Y = \beta) \\
&= \sum_\alpha \sum_\beta \alpha\beta \Pr(X = \alpha) \Pr(Y = \beta) \\
&= \left(\sum_\alpha \alpha \Pr(X = \alpha) \right) \left(\sum_\beta \beta \Pr(Y = \beta \right) \\
&= E(X)E(Y)
\end{aligned}$$

□

This is *not* necessarily true if X and Y are not independent.

3 Useful Inequalities

3.1 Markov Inequality

Theorem 3.1. *Let X be a random variable over Ω so that X can never take a negative value. Then $P(X \geq t) \leq E(X)/t$ if $t \geq 1$.*

Proof.

$$\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} k \Pr(X = k) \\
&\geq \sum_{k=t}^{\infty} k \Pr(X = k) \\
&\geq \sum_{k=t}^{\infty} t \Pr(X = k) \\
&= t \Pr(X \geq t)
\end{aligned}$$

□

This implies that $\Pr(X \neq 0) \leq E(X)$.

An example is n distinguishable balls, m boxes. Z is number of empty boxes. Then we can compute $E(Z)$ fairly easily, thus giving a bound on the probability that a box is empty.

4 Variance

We have talked about expectation – the "average" value of a random variable – the value that we would "expect" the variable to take. Also, we would like to know just how "reliable" our estimate is. Variance gives us a way of determining just that:

Definition 4.1 (Variance). *Let $Z : \Omega \to \mathbb{R}$. Let $E(Z) = \mu$. Then $Z - \mu$ is a random variable, and we define the variance of Z, denoted* $\mathrm{Var}(Z)$ *to be* $E\left((Z-\mu)^2\right)$.

Therefore, the variance is always nonnegative, and in any interesting case is positive.

Fact 4.2.

$$\mathrm{Var}(Z) = E(Z^2) - (E(Z))^2$$

Proof.

$$\begin{aligned}
\mathrm{Var}(Z) &= E((Z-\mu)^2) \\
&= E(Z^2 - 2\mu Z + \mu^2 \\
&= E(Z^2) - 2\mu E(Z) + \mu^2 \\
&= E(Z^2) - \mu^2
\end{aligned}$$

□

What is the variance in rolling two dice, if Z is the sum of the two numbers on top?

4.1 Binomial Distribution

Let $Z + B_{n,p}$. Then, $E(Z) = \mu = np$.

What is the variance?

$$
\begin{aligned}
\mathrm{Var}(B_{n,p}) &= \left(\sum_{k=1}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k}\right) - \mu^2 \\
&= \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} + \mu - \mu^2 \\
&= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2}(1-p)^{n-k} + \mu - \mu^2 \\
&= n(n-1)p^2 (p + (1-p))^{n-2} + \mu - \mu^2 \\
&= n(n-1)p^2 + \mu - \mu^2 \\
&= np(1-p)
\end{aligned}
$$

We can use variance to bound the probability that a variable is too far from its mean.

Let $\sigma = \sqrt{\mathrm{Var}(Z)}$. Then we have:

$$
\begin{aligned}
\Pr(|Z-\mu| \geq t\sigma) &= \Pr((Z-\mu)^2 \geq t^2\sigma^2) \\
&\leq \frac{E((Z-\mu)^2)}{t^2\sigma^2} \\
&= \frac{1}{t^2}
\end{aligned}
$$

The middle line comes from the Markov inequality.

What does this tell us about the binomial distribution?

Well, in the binomial case, $\sigma = \sqrt{np(1-p)}$.

So $\Pr(|B_{n,p} - np| \geq t\sigma) \leq \frac{1}{t^2}$.

Therefore, $\Pr(|B_{n,p} - np| \geq \epsilon np) \leq \frac{1}{\epsilon^2 np}$.

This is the law of large numbers.