COS 341, October 12, 2001
Handout Number 5

Some Probability Theory

1. Expected Value and Variance

Let X be a random variable on a probability space $\Omega = (U, p)$. The *expected value* of X is defined as $E(X) = \sum_{u \in U} p(u)X(u)$, and the *variance* of X is defined as $Var(X) = \sum_{u \in U} p(u)(X(u) - E(X))^2$. Clearly $Var(X) \geq 0$. The *standard deviation* of X is defined as $\sigma(X) = \sqrt{Var(X)}$. Intuitively, the value $E(X)$ tells you what value of X you can expect, if you perform many many experiments; the variance gives you some idea on how close to $E(X)$ you expect to see the typical value of X fall in these experiments. This last statement is made more precise in the Chebyshev's Inequality in the future.

As an illustration, let $\Omega = (U, p)$, where $U = \{a, b, c, d, e, f, g, h\}$ and $p(a) = 0.2, p(b) = 0.13, p(c) = 0.17, p(d) = 0.1, p(e) = 0.1, p(f) = 0.2, p(g) = 0.04, p(h) = 0.06$. Let X be the random variable on Ω with $X(a) = 1, X(b) = X(c) = 3, X(d) = X(e) = X(f) = 4, X(g) = X(h) = 9$. Then by defintion

$$E(X) = p(a)X(a) + p(b)X(b) + \cdots + p(h)X(h).$$

Let $t = E(X)$ denote the above value. Then by definition

$$Var(X) = p(a)(X(a) - t)^2 + p(b)(X(b) - t)^2 + \cdots + p(h)(X(h) - t)^2.$$

The numerical values of $E(X)$ and $Var(X)$ can be calculated straightforwardly from these formulae. We shall not do the calculation here. The results are $E(X) = 3.6$, $Var(X) = 4.440$, and $\sigma(X) = \sqrt{4.440} = 2.1\cdots$.

If c is a real number and X is a random variable, then cX denotes the random variable Y such that $Y(u) = cX(u)$ for all $u \in U$. If $X_i, 1 \leq i \leq m$ are random variables in Ω, then the *sum* of the random variables X_i is the random variable $W = \sum_{1 \leq i \leq m} X_i$, defined by $W(u) = \sum_{1 \leq i \leq m} X_i(u)$ for all $u \in U$. The *product* of the random variables X_i is the random variable $Z = \prod_{1 \leq i \leq m} X_i$, defined by $Z(u) = \prod_{1 \leq i \leq m} X_i(u)$ for all $u \in U$.

Note that we have $Var(X) = E(Z)$, where $Z = (X - E(X))^2$. (Keep in mind that $E(X)$ is a real number t, and for all $u \in U$, the random variable $X - E(X)$ has value $X(u) - t$, the random variable Z has value $Z(u) = (X(u) - t)^2$.)

If $X_i, 1 \leq i \leq m$ are random variables in Ω and c_i are real numbers. Let $W = \sum_{1 \leq i \leq m} c_i X_i$.

Useful Formula 1: Linearity of the Expected Value

$$E(W) = \sum_{1\le i\le m} c_i E(X_i).$$

Proof

$$\begin{aligned} E(W) &= \sum_{u\in U} p(u)W(u) \\ &= \sum_{u\in U} p(u) \sum_{1\le i\le m} c_i X_i(u) \\ &= \sum_{1\le i\le m} c_i \sum_{u\in U} p(u)X_i(u) \\ &= \sum_{1\le i\le m} c_i E(X_i). \end{aligned}$$

This proves the formula. □

Useful formula 2

$$Var(X) = E(X^2) - (E(X))^2. \tag{1}$$

Proof Let $t = E(X)$ and $Z = (X-t)^2$. Note that $Z(u) = (X-t)^2(u) = (X(u))^2 - 2tX(u) + t^2$ for all $u \in U$. Thus $(X-t)^2 = X^2 - 2tX + t^2$ expresses the random variable Z as a linear combination of the random variables X^2, X. By the linearity of the expected value, we have $E(Z) = E(X^2) - 2tE(X) + t^2 = E(X^2) - t^2$, which proves (1). □

2. Generating Functions

We have seen earlier that generating functions are useful for evaluating sums such as $\sum_{k=0,2,4,\cdots} \binom{n}{k}$. We now demonstrate that the generating functions are also useful for evaluating $E(X)$ and $Var(X)$, if X takes on only nonnegative integer values. Take any such X, let

$$a_k = \Pr\{X = k\} = \sum_{\substack{u\in U \\ X(u)=k}} p(u).$$

Define

$$A(x) = \sum_{k\ge 0} a_k x^k.$$

Then

$$A'(x) = \sum_{k\ge 0} k a_k x^{k-1},$$

and

$$A''(x) = \sum_{k\ge 0} k(k-1) a_k x^{k-2}.$$

Thus,

$$A'(1) = \sum_{k\geq 0} k a_k,$$

and

$$A''(1) = \sum_{k\geq 0} k(k-1) a_k.$$

It follows that

$$E(X) = \sum_{k\geq 0} k a_k = A'(1). \tag{2}$$

And

$$\begin{aligned} E(X^2) &= \sum_{k\geq 0} k^2 a_k \\ &= \sum_{k\geq 0} (k(k-1)+k) a_k \\ &= \sum_{k\geq 0} k(k-1)a_k + \sum_{k\geq 0} k a_k \\ &= A''(1) + A'(1). \end{aligned}$$

By (1) this gives

$$Var(X) = A''(1) + A'(1) - (A'(1))^2. \tag{3}$$

As an application, consider the example of tossing a fair coins n times, and let X be the number of Heads in the sequence. Then $A_k = \binom{n}{k}/2^n$ and $A(x) = \sum_k a_k x^k = (1+x)^n/2^n$. Clearly, $A'(x) = n(1+x)^{n-1}/2^n$ and $A''(x) = n(n-1)(1+x)^{n-2}/2^n$. Thus $A'(1) = n/2$ and $F''(1) = n(n-1)/4$. From (1) we have $Var(X) = n(n-1)/4 + (n/2) - (n^2/4) = n/4$.