COS 341, October 6, 1997
Handout Number 7

Some Probability Theory

Let X be a random variable on a probability space $\Omega = (U, p)$. We define $E(X) = \sum_{u \in U} p(u)X(u)$, and $Var(X) = E(Z)$, where $Z = (X - E(X))^2$. (Note 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 $Z(u) = (X(u) - t)^2$. Clearly $Var(X) \geq 0$.) The *standard deviation* $\sigma(X)$ is defined to be $\sqrt{Var(X)}$. Intuitively, the value $E(X)$ tells you what the average 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 below.

Recall that, if $X_i, 1 \leq i \leq m$ are random variables in Ω and c_i are real numbers, then the random variable $W = \sum_{1 \leq i \leq m} c_i X_i$ is defined by $W(u) = \sum_{1 \leq i \leq m} c_i X_i(u)$. The following important formula was derived in class.

Linearity of Expected Value

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

As an example to illustrate all these discussions, 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. Rather, we calculate them in a slightly different way.

Clearly, $\Pr\{X = 1\} = p(a) = 0.2$, $\Pr\{X = 3\} = p(b) + p(c) = 0.3$, $\Pr\{X = 4\} = p(d) + p(e) + p(f) = 0.4$, and $\Pr\{X = 9\} = p(g) + p(h) = 0.1$.

Now, by definition, $E(X) = \sum_{u \in U} p(u)X(u)$. When X takes on only nonnegative integer values, we can rewrite it as

$$E(X) = \sum_{k \geq 0} \Pr\{X = k\}k. \tag{1}$$

Similarly, from the definition of $Var(X)$, we can derive

$$Var(X) = \sum_{k \geq 0} \Pr\{X = k\}(k - E(X))^2. \tag{2}$$

Thus, for the example above,

$$E(X) = 0.2 \cdot 1 + 0.3 \cdot 3 + 0.4 \cdot 4 + 0.1 \cdot 9 = 3.6,$$

and

$$Var(X) = 0.2 \cdot (1 - 3.6)^2 + 0.3 \cdot (3 - 3.6)^2 + 0.4 \cdot (4 - 3.6)^2 + 0.1 \cdot (9 - 3.6)^2 = 4.440.$$

Also we have $\sigma(X) = \sqrt{4.440} = 2.1 \cdots$.

The following inequality says that, taking an experiment according to Ω, the probability of seeing the value of X deviating from $E(X)$ by c (or more) standard deviations is less than $1/c^2$.

Chebyshev's Inequality Let $c > 0$,

$$\Pr\{|X - E(X)| > c\sigma(X)\} \leq c^{-2}.$$

Proof If $Var(X) = 0$, then $X(u) = E(X)$ for all $u \in U$ (why?), and the Inequality is clearly true.

Assume $Var(X) > 0$. Let $Y = X - E(X)$. Then

$$\begin{aligned}
Var(X) &= E(Y^2) \\
&= \sum_{u \in U} p(u)(Y(u))^2 \\
&\geq \sum_{u \in U, |Y(u)| > c\sigma(X)} p(u)(Y(u))^2 \\
&\geq (c\sigma(X))^2 \sum_{u \in U, |Y(u)| > c\sigma(X)} p(u) \\
&= c^2 Var(X) \Pr\{|Y(u)| > c\sigma(X)\}.
\end{aligned}$$

This implies the Inequality. $\square$

For example, take $c = 10$. The probability for X to deviate from $E(X)$ by $10\sigma(X)$ (or more) is less than 1 percent.

Another useful formula

$$Var(X) = E(X^2) - (E(X))^2. \tag{3}$$

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 (3). $\square$

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 $Var(X)$, when X take on only nonnegative integer values. Take any such X, let $p_k = \Pr\{X = k\}$. Define

$$F(x) = \sum_{k \geq 0} p_k x^k.$$

Then

$$F'(x) = \sum_{k \geq 0} k p_k x^{k-1},$$

and

$$F''(x) = \sum_{k \geq 0} k(k-1) p_k x^{k-2}.$$

Thus,

$$F'(1) = \sum_{k \geq 0} k p_k,$$

and

$$F''(1) = \sum_{k \geq 0} k(k-1) p_k.$$

It follows that

$$E(X) = \sum_{k \geq 0} k p_k = F'(1). \tag{4}$$

And

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

By (3) this gives

$$Var(X) = F''(1) + F'(1) - (F'(1))^2. \tag{5}$$

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 $p_k = \binom{n}{k}/2^n$ and $F(x) = \sum_k p_k x^k = (1+x)^n/2^n$. Clearly, $F'(x) = n(1+x)^{n-1}/2^n$ and $F''(x) = n(n-1)(1+x)^{n-2}/2^n$. Thus $F'(1) = n/2$ and $F''(1) = n(n-1)/4$. From (5) we have $Var(X) = n(n-1)/4 + (n/2) - (n^2/4) = n/4$.