Lecture 4: Discrete Probability Distributions

(Text Sections 2.2, 2.4)

Bernoulli Distribution

If a random variable X can take on only two values (labelled 0="failure" and 1="success"), then the distribution of X is, by definition, Bernoulli(p), where p is the probability of success. The pmf of X can be written as

$$p(x) = P(X = x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases} .$$

Example: Say a patient (unknowingly) has strep throat. The doctor uses a throat swab test, which comes back negative with probability $1 - p$ or positive with probability p. Defining correct diagnosis as success, the outcome of the test has a Bernoulli(p) distribution. In this context, the probability p is known as the *sensitivity* of the test.

Binomial Distribution

If we observe n independent trials, each of which results in "success" with probability p and "failure" with probability $1 - p$, then the number of successes, X has, by definition, a Binomial(n, p) distribution. Specifically, the pmf of X is

$$p(x) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

Example: If n patients with strep throat come to the ER, then the number who receive a positive (correct) diagnosis is distributed as Binomial(p) (assuming that test results are independent, e.g. swabs are not processed in batches). The probability that at least k of these n people receive a positive diagnosis is

$$P(X \geq k) = \sum_{x=k}^{n} p(x).$$

Geometric Distribution

If we again observe n independent trials, each of which results in "success" with probability p and "failure" with probability $1 - p$, then the number of *trials* required to observe the first success, X has, by definition, a Geometric(p) distribution. Specifically, the pmf of X is

$$p(x) = P(X = x) = (1 - p)^{x-1} p, \qquad x = 1, 2, \ldots.$$

NOTE: Alternatively, we might say that the number of *failures* observed before observing the first success, Y, has a Geometric(p) distribution, i.e.

$$p(y) = P(Y = y) = (1 - p)^y p, \qquad y = 0, 1, \ldots.$$

The two specifications are effectively the same, but the definition of the random variable and its support are critical in interpreting them correctly!

Example: If a stream of patients with strep throat come to the ER, then the number required to test before observing one positive test result has a Geometric(p) distribution (assuming test results are independent).

NOTE: The geometric distribution is the only *memoryless* discrete distribution: for $y < x$,

$$\begin{aligned}
P(X = x \mid X \geq y) &= \frac{P(X = x, X \geq y)}{P(X \geq y)} \\
&= \frac{P(X = x)}{P(X \geq y)} \\
&= \frac{(1-p)^{x-1}p}{\sum_{x=y}^{\infty}(1-p)^{x}p} \\
&= \frac{(1-p)^{x-1}}{\frac{(1-p)^{y}}{p}} \\
&= (1-p)^{x-y-1}p \\
&= P(X = x - y).
\end{aligned}$$

In other words, the number of failed trials that have already occurred does not affect the distribution of the remaining number of trials required to observe the first success.

Poisson Distribution

A random variable X taking on values in $\{0, 1, 2, \ldots\}$ has a Poisson(μ) distribution if

$$p(x) = P(X = x) = \frac{e^{-\mu}\mu^{x}}{x!}, \qquad x = 0, 1, 2, \ldots$$

where $\mu > 0$.

Example: Let X be the number of times in a year a BCAA member calls for roadside assistance. It might be reasonable to assume that X has a Poisson(μ) distribution. Knowledge of μ can help BCAA to establish annual premiums for this member.

NOTE: One critical difference between the binomial and Poisson distributions is that the support of the binomial distribution is $[0, n]$, for $n < \infty$, while the support of the Poisson distribution is $[0, \infty)$. However, the binomial distribution becomes closer and closer to the Poisson distribution as $n \to \infty$ (see text Section 2.2.4). Another difference is that the binomial distribution is the *only* distribution that can describe the number of successes in n independent trials where each has the same probability of success. In contrast, there are many distributions that can describe unbounded counts.

Discrete Uniform Distribution

A random variable X taking on values in $\{m, m+1, \ldots, n\}$ has a Uniform$[m, n]$ distribution if

$$p(x) = P(X = x) = 1/(n - m + 1), \qquad x = m, m + 1, \ldots, n.$$

Example: The outcome of the roll of a fair die, X, has a Uniform$[1, 6]$ distribution, since $P(X = x) = 1/6$, $x = 1, \ldots, 6$.

Expectation of a Discrete Random Variable

Definition: The *expected value*, *expectation*, or *mean* of a discrete random variable X is defined by

$$\mathrm{E}[X] = \sum_x x p(x).$$

Thus, $\mathrm{E}[X]$ is a weighted average of all the values that X can take on. It is a measure of central tendency. It is not, however, an indication of the "most likely value" that a random variable can take on!

Example: Consider the outcome of the roll of a fair die, X.

$$\mathrm{E}[X] = \sum_{x=1}^{6} x(1/6) = 3.5.$$

$\mathrm{E}[X]$ is an indication of the central tendency of X in the sense that it is an average of the values that X can take on, but $P(X = \mathrm{E}[X]) = 0$ in this case!

Example: Consider the population of (very bright!) SFU students who have (rounded) GPAs of 3.7, 3.8, 3.9, or 4.0. Let the GPA of a student within this group be X. Say $P(X = 3.7) = 0.5$, $P(X = 3.8) = 0.3$, $P(X = 3.9) = 0.1$, and $P(X = 4.0) = 0.1$. Note that X is discrete even though it takes on non-integer values! Then

$$\mathrm{E}[X] = 3.7(0.5) + 3.8(0.3) + 3.9(0.1) + 4.0(0.1) = 3.78.$$

Example: Let $X \sim$Poisson(μ). Then

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{x=0}^{\infty} x \cdot \frac{e^{-\mu}\mu^x}{x!} \\
&= \sum_{x=1}^{\infty} x \cdot \frac{e^{-\mu}\mu^x}{x!} \\
&= \mu \sum_{x=1}^{\infty} \frac{e^{-\mu}\mu^{x-1}}{(x-1)!} \\
&= \mu \sum_{y=0}^{\infty} \frac{e^{-\mu}\mu^{y}}{y!} \\
&= \mu
\end{aligned}$$

Thus, the parameter μ of a Poisson(μ) distributed random variable represents its mean.

Returning to our BCAA example, if $\mu = 2.4$ and each call costs BCAA on average \$30, then the annual premium for this member would have to be at least 2.4(\$30)=\$72.00 in order for BCAA to make a profit, on average.

Expectation of a Function of a Discrete Random Variable

Recall that a random variable X is simply a real-valued function defined on the sample space, S. Thus, any real-valued function $g(X)$ is also a random variable.

Example: Let X be the outcome of the roll of a fair die, so that $X \in \{1, \ldots, 6\}$. Let g be an indicator variable, and define $Y = 1\{X \leq 3\}$. Then Y is a real-valued function of X, and hence is also a random variable. Here, $Y = 1$ if the outcome is ≤ 3, and $Y = 0$ otherwise.

We can use this idea to compute expectations of functions of random variables. Specifically,

$$\mathrm{E}[g(X)] = \sum_x g(x)p(x).$$

One particularly important example is the *variance* of a random variable. Letting $\mu = \mathrm{E}[X]$,

$$\mathrm{Var}[X] = \mathrm{E}[(X-\mu)^2] = \sum_x (x-\mu)^2 p(x).$$

$$\text{Also, } \mathrm{Var}[X] = \mathrm{E}[X^2] - \mu^2 = \sum_x x^2 p(x) - \mu^2.$$

Example: Compute the variance of a random variable with Poisson(μ) distribution.

$$\begin{aligned}
\mathrm{Var}[X] &= \sum_{x=0}^{\infty} x^2 \cdot \frac{e^{-\mu}\mu^x}{x!} - \mu^2 \\
&= \sum_{x=1}^{\infty} \frac{xe^{-\mu}\mu^x}{(x-1)!} - \mu^2 \\
&= \sum_{x=1}^{\infty} \left[\frac{xe^{-\mu}\mu^x}{(x-1)!} - \frac{e^{-\mu}\mu^x}{(x-1)!} + \frac{e^{-\mu}\mu^x}{(x-1)!}\right] - \mu^2 \\
&= \sum_{x=1}^{\infty} \left[\frac{(x-1)e^{-\mu}\mu^x}{(x-1)!} + \frac{e^{-\mu}\mu^x}{(x-1)!}\right] - \mu^2 \\
&= \sum_{x=2}^{\infty} \frac{(x-1)e^{-\mu}\mu^x}{(x-1)!} + \sum_{x=1}^{\infty} \frac{e^{-\mu}\mu^x}{(x-1)!} - \mu^2 \\
&= \sum_{x=2}^{\infty} \frac{e^{-\mu}\mu^x}{(x-2)!} + \mu \sum_{y=0}^{\infty} \frac{e^{-\mu}\mu^y}{y!} - \mu^2 \\
&= \mu^2 \sum_{y=0}^{\infty} \frac{e^{-\mu}\mu^y}{y!} + \mu - \mu^2 \\
&= \mu^2 + \mu - \mu^2 \\
&= \mu
\end{aligned}$$

So, the mean and variance of a Poisson distributed random variable are equal.