Bernoulli Trials and the Binomial and Geometric Distributions

A random variable that assumes only the values 0 and 1 is called a Bernoulli variable. An experiment with only two outcomes is called a Bernoulli trial. In particular, if an experiment can only result in a "success" or "failure", then it is a Bernoulli trial. We often associate a 1 with a "success" and a 0 with a "failure".

If X is a Bernoulli variable, then $E(X) = p$, the probability of "success". To see this, consider

$$E(X) = \sum_{i=1}^{2} x_i p(x_i) = 1(p) + 0(1-p) = p\,.$$

The variance of X is

$$\begin{aligned} V(X) &= \sum_{i=1}^{2} (x_i - \mu)^2 \, p(x_i) \\ &= (1-p)^2 \, p + (0-p)^2 (1-p) \\ &= (1-p)\, p\,(p + 1 - p) \\ &= (1-p)\, p, \end{aligned}$$

so the standard deviation of X is $\sigma_X = \sqrt{(1-p)\,p}$.

The Binomial Distribution

Repeated independent Bernoulli trials results in the binomial distribution when the quantity of interest is the number of successes in a certain number of trials. The probability of success remains constant between trials. If we repeat a Bernoulli experiment n times, we have the distribution $B(n, p)$.

The mean of the binomial distribution $B(n, p)$ can be found, since we know that the expected value of a sum is the sum of the expected values. So,

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = np\,.$$

In a similar way, since the trials are independent, we know the variance of the sum is the sum of the variances. So, the variance of $B(n, p)$ is

$$V(X_1 + X_2 + \cdots + X_n) = V(X_1) + V(X_2) + \cdots + V(X_n) = n\big(p(1-p)\big)\,.$$

Consequently, the standard deviation of the binomial distribution $B(n, p)$ is $\sqrt{np(1-p)}$.

So, if $X \sim B(n, p)$, then $E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = np$ and

The Geometric Distribution

If we have repeated Bernoulli trials each independent with probability of success *p*, but our random variable *X* is the trial on which the first success occurs, then we have a geometric distribution, denoted *G*(*p*). The pdf for *G*(*p*) is $P\left(x=k\right)=p\left(1-p\right)^{k-1}$.

So, $E\left(X\right)=\sum_{k=1}^{\infty}k\left(p\left(1-p\right)^{k-1}\right)=p\sum_{k=1}^{\infty}k\left(1-p\right)^{k-1}$. This sum is easy to find if we notice that $k\left(1-p\right)^{k-1}$ is the opposite of the derivative of $\left(1-p\right)^{k}$ with respect to *p*. So, $p\sum_{k=1}^{\infty}k\left(1-p\right)^{k-1}=-p\sum_{k=1}^{\infty}\frac{d}{dp}\left(1-p\right)^{k}$. Since the derivative of a sum is the sum of the derivatives (having an infinite sum can create problems here, but in this case it does not), so $p\sum_{k=1}^{\infty}k\left(1-p\right)^{k-1}=-p\frac{d}{dp}\left(\sum_{k=1}^{\infty}\left(1-p\right)^{k}\right)$. This infinite sum is now geometric, so we can compute it easily. We have $-p\frac{d}{dp}\left(\frac{1-p}{1-\left(1-p\right)}\right)=-p\frac{d}{dp}\left(\frac{1}{p}-1\right)=-p\left(\frac{-1}{p^{2}}\right)=\frac{1}{p}$.

So, we have if $X\sim G\left(p\right)$, then $E\left(X\right)=\frac{1}{p}$.

We can compute the variance of X using the computational formula,

$$Var\left(X\right)=E\left(X^{2}\right)-E\left(X\right)^{2}.$$

If we can compute $E\left(X^{2}\right)=\sum_{k=1}^{\infty}k^{2}p\left(1-p\right)^{k-1}$, we will have all we need. To simplify, let $q=\left(1-p\right)$ and focus on the summation $\sum_{k=1}^{\infty}k^{2}q^{k-1}=1+4q+9q^{2}+16q^{3}+25q^{4}+\cdots$. How can we create this sum from a geometric series, which is the only series we are sure we can compute.

We know that $1+q+q^{2}+q^{3}+q^{4}+\cdots=\frac{1}{1-q}$ when $\left|q\right|<1$. By differentiating both sides of this equation, we have $1+2q+3q^{2}+4q^{3}+5q^{4}+\cdots=\frac{1}{\left(1-q\right)^{2}}$ when $\left|q\right|<1$. Multiplying both sides by *q*, give us $q+2q^{2}+3q^{3}+4q^{4}+5q^{5}+\cdots=\frac{q}{\left(1-q\right)^{2}}$ when $\left|q\right|<1$. Now, if we differentiate again with respect to *q* using the quotient rule, we have $1+4q+9q^{2}+16q^{3}+25q^{4}+\cdots=\frac{1+q}{\left(1-q\right)^{3}}$ when $\left|q\right|<1$. Since $q=1-p<1$, the sum of $1+4q+9q^{2}+16q^{3}+25q^{4}+\cdots$ is $\frac{2-p}{p^{3}}$.

Now, substitute back to find that $E\left(X^2\right) = p\sum_{k=1}^{\infty} k^2 (1-p)^{k-1} = p\left(\frac{2-p}{p^3}\right) = \frac{(2-p)}{p^2}$.

So, $Var(X) = \frac{(2-p)}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}$.

So, if $X \sim G(p)$, then $E(X) = \frac{1}{p}$ and $Var(X) = \frac{1-p}{p^2}$.