**Stat 381: Applied Statistical Methods I** **Spring 2015**

February 18, 20

*Instructor: Brian Powers*

# 15.1 Discrete Probability Distributions

## 15.1.1 Uniform Distribution

**Definition 15.1.** A random variable $X$ which can take any integer value from $a$ to $b$ (inclusive) with equal probability is said to follow a **discrete uniform distribution**. As notation we say $X \sim Unif_D(a, b)$.

**Theorem 15.2.** *If $X \sim Unif_D(a, b)$, its pmf is given as*

$$f(x) = \frac{1}{b-a+1} \quad x \in \{a, \ldots, b\}$$

*Its expected value and variance are*

$$E(X) = \frac{a+b}{2}, Var(X) = \frac{(b-a+1)^2+1}{12}.$$

*Proof.* There are $b - a + 1$ values that $X$ can take, thus the pmf.

$$E(X) = \sum_{x=a}^{b} x\frac{1}{b-a+1} = \frac{1}{b-a+1}\left(\sum_{x=1}^{b} x - \sum_{x=1}^{a-1} x\right) = \frac{1}{b-a+1}\left(\frac{b(b+1)}{2} - \frac{(a-1)(a)}{2}\right) = \frac{b+a}{2}$$

The proof for variance is omitted, but it is proved directly using summations. □

## 15.1.2 Bernoulli Distribution

**Definition 15.3.** A random variable $X$ which can take values 1 (a success) or 0 (a failure) with respective probabilities $p$ and $q = 1 - p$ follows a **Bernoulli distribution**, and we denote this as $X \sim Bern(p)$.

**Theorem 15.4.** *If $X \sim Bern(p)$,*

$$E(X) = p, Var(X) = pq.$$

*Proof.*

$$E(X) = \sum_x xf(x) = 0(q) + 1(p) = p$$

$$E(X^2) = \sum_x x^2 f(x) = 0^2(q) + 1^2(p) = p$$

$$Var(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p) = pq$$

□

### 15.1.3 Binomial Distribution

**Definition 15.5.** Suppose independent random variables $X_1, X_2, \dots, X_n \sim Bern(p)$. Let $Y = \sum_{i=1}^{n} X_i$, the count of how many successes are achieved after $n$ Bernoulli trials. $Y$ follows a **binomial distribution** with parameters $n$ and $p$, denoted $Y \sim Binom(n, p)$.

**Theorem 15.6.** *If $Y \sim Binom(n, p)$ then the pmf is given by*

$$f(y) = \binom{n}{y} p^y (1-p)^{n-y} \quad y \in \{0, \dots, n\}$$

*The expected value and variance are*

$$E(Y) = np, Var(Y) = npq.$$

*Proof.* $f(y) = P(Y = y)$, that is, the probability that exactly $y$ of the $n$ trials are successes. Suppose the first $y$ trials were successes and the last $n - y$ were failures. Because the trials are all independent of each other, the probability that this would occur is $p^y(1-p)^{n-x}$. However, any $y$ of the $n$ trials could have been the successes, and any choice of $y$ trials is equally likely. So $P(X = y) = \binom{n}{y} p^x (1-p)^{n-y}$.
Since $Y = \sum_{i=1}^{n} X_i$,

$$E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np$$

$$Var(Y) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} pq = npq$$

□

### 15.1.4 Hypergeometric Distribution

**Definition 15.7.** If a population of size $N$ consists of $k$ individuals with a certain characteristic and $N - k$ individuals without the characteristic, and we take a sample of $n$ without replacement, let $X$ be the number of individuals in the sample with the characteristic in question. $X$ follows a **hypergeometric distribution** with parameters $N, k, n$, denoted $X \sim Hypergeom(N, k, n)$.

**Theorem 15.8.** *If $X \sim Hypergeom(N, k, n)$, the pmf is given by*

$$f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \quad x = \max\{0, n+k-N\}, \dots, \min\{n, k\}$$

*The expected value and variance are*

$$E(X) = \frac{nk}{N}, Var(X) = \frac{nk(N-k)(N-n)}{N^2(N-1)}$$

The proof is omitted. The pmf can be justified by reasoning, it is more difficult and beyond the scope of this course to prove the expected value or variance.

### 15.1.5 The Negative Binomial Distribution

**Definition 15.9.** Consider a Bernoulli process with probability $p$ generating independent random variables $X_1, X_2, \ldots$. Let $Y$ be the number of trials required to produce $k$ successes. The random variable $Y$ follows a **negative binomial distribution** with parameters $p$ and $k$. We denote this as $Y \sim NegBinom(k, p)$.

**Theorem 15.10.** *Suppose $Y \sim NegBinom(k, p)$. The pmf of $Y$ is*

$$f(y) = \binom{y-1}{k-1} p^k q^{y-k}, \quad y = k, k+1, \ldots$$

*Proof.* $f(y) = P(Y = y)$ that is, it takes $y$ trials to produce $k$ successes. This means that trial $y$ was a success, and among the $y-1$ prior trials, $k-1$ of them were successes while $y-k$ of them were failures. The probability that $k$ successes and $y-k$ failures would occur (since each trial is independent) is $p^k q^{y-k}$, and since this could happen in $\binom{y-1}{k-1}$ ways, the pmf is derived. □

### 15.1.6 Geometric Distribution

**Definition 15.11.** Consider a Bernoulli process with probability $p$ generating independent random variables $X_1, X_2, \ldots$. Let $Y$ be the number of trials until the first success. Random variable $Y$ follows a **geometric distribution** with parameter $p$, denoted $Y \sim Geom(p)$.

**Theorem 15.12.** *Suppose $Y \sim Geom(p)$. The pmf is given by*

$$f(y) = pq^{y-1}, \quad y = 1, 2, \ldots$$

*The expected value and variance are*

$$E(Y) = \frac{1}{p}, Var(Y) = \frac{1-p}{p^2}.$$

*Proof.* $f(y) = P(Y = y)$ that is, the probability that the first $y-1$ trials are failures and trial $y$ is a success. Since the trials are independent, the probability that this should occur is $pq^{y-1}$.

$$E(Y) = \sum_{y=1}^{\infty} ypq^{y-1} = \frac{p}{q} \sum_{y=1}^{\infty} yq^y = \frac{p}{q}\left(\frac{q}{(1-q)^2}\right) = \frac{1}{p}$$

The proof for variance is omitted and is beyond the scope of this course. □

### 15.1.7 Poisson Distribution

Now we consider the number of times some "rare" event occurs during a given time interval. We consider random phenomena that have the following properties:

1. **Memoryless:** The number of times the event occurs during any time interval is independent of how many times it occurs during any other disjoint time interval.
2. The probability that a single event occurs during a time interval is proportional to the length of the time interval and is independent of how many events occur outside of this interval.
3. The probability that more than one event occurs simultaneously is zero.

**Definition 15.13.** Any process which generates events following properties 1,2, and 3 above is called a **Poisson process**. If the number of times an event occurs during a fixed time period is $X$, we say $X$ follows a **Poisson distribution**. If $\lambda$ is the average number of events occuring during 1 unit of time, and $X$ is counted during $t$ units of time, we say $X \sim Poisson(\lambda t)$.

**Theorem 15.14.** *If $X \sim Poisson(\lambda t)$, the pmf of $X$ is*

$$f(x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0, 1, 2, \ldots$$

The proof is beyond this course.

**Theorem 15.15.** *Suppose random variable $X \sim Poisson(\lambda t)$. Then*

$$E(X) = \lambda t, Var(X) = \lambda t.$$

*Proof.* Suppose $X \sim Poisson(\lambda t)$

$$\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda t}(\lambda t)^x}{x!} \\
&= e^{-\lambda t} \sum_{x=0}^{\infty} x \frac{(\lambda t)^x}{x!} \\
&= e^{-\lambda t} \sum_{x=1}^{\infty} x \frac{(\lambda t)^x}{x(x-1)!} && \text{first term is 0} \\
&= \lambda t e^{-\lambda t} \sum_{x=1}^{\infty} \frac{(\lambda t)^{x-1}}{(x-1)!} \\
&= \lambda t e^{-\lambda t} \sum_{x=0}^{\infty} \frac{(\lambda t)^x}{x!} && \text{reindex starting at 0} \\
&= \lambda t e^{-\lambda t} \left(e^{\lambda t}\right) && \text{series summation for } e^{\lambda t} \\
&= \lambda t
\end{aligned}$$

The proof for variance is omitted. □