Some common families of discrete random variables

Math 30530, Fall 2012

October 7, 2012

The Bernoulli random variable

- **Name**: Bernoulli(p)
- **When to use**: When you want to indicate whether an experiment resulted in success or not; Bernoulli random variable takes value 1 if success occurred, and 0 otherwise
- **Parameter**:
  - p: the probability of success (so $p = \Pr(A)$ if success is that event A occurred)
- **Mass function**:

$$p_X(x) = \begin{cases} p & \text{if } x = 1 \\ q = 1 - p & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

- **Statistics**:
  - $\mu = E(X) = p$
  - $\sigma^2 = \mathrm{Var}(X) = p(1 - p) = pq$

The Binomial random variable

- **Name**: $\mathrm{Binomial}(n, p)$
- **When to use**: When you want to count how many successes you had, when you repeat the **same** experiment a **fixed** number of times, **independently** of each other
- **Parameters**:
  - n: the number of times the experiment is repeated
  - p: the probability of success on each individual trial
- **Mass function**:

$$p_X(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & \text{if } x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

  where $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ counts the number of ways of distributing x successes among n trials, and $n! = n \times (n-1) \times \ldots \times 3 \times 2 \times 1$
- **Statistics**:
  - $\mu = E(X) = np$
  - $\sigma^2 = \mathrm{Var}(X) = npq$

The Geometric random variable

- **Name**: $\text{Geometric}(p)$
- **When to use**: When you want to count how many times you have to repeat the **same** experiment, **independently** of each other, until you first have success
- **Parameter**:
  - p: the probability of success on each individual trial
- **Mass function**:

$$p_X(x) = \begin{cases} q^{x-1}p & \text{if } x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

- **Statistics**:
  - $\mu = E(X) = \frac{1}{p}$
  - $\sigma^2 = \text{Var}(X) = \frac{q}{p^2}$

The Negative Binomial random variable

- **Name**: $\mathrm{NegBinomial}(r, p)$
- **When to use**: When you want to count how many times you have to repeat the **same** experiment, **independently** of each other, until you first have some predetermined number of successes
- **Parameters**:
  - r: the number of successes you are aiming for
  - p: the probability of success on each individual trial
- **Mass function**:

$$p_X(x) = \begin{cases} \binom{x-1}{r-1} q^{x-r} p^r & \text{if } x = r, r+1, r+2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

- **Statistics**:
  - $\mu = E(X) = \frac{r}{p}$
  - $\sigma^2 = \mathrm{Var}(X) = \frac{rq}{p^2}$

The Discrete Uniform random variable

- **Name**: $\mathrm{D.Uniform}(N)$
- **When to use**: When you are assigning values 1 through N to N equally likely outcomes
- **Parameter**:
  - N: the number of outcomes
- **Mass function**:

$$p_X(x) = \begin{cases} \frac{1}{N} & \text{if } x = 1, 2, 3, \ldots, N \\ 0 & \text{otherwise} \end{cases}$$

- **Statistics**:
  - $\mu = E(X) = \frac{N+1}{2}$
  - $\sigma^2 = \mathrm{Var}(X) = \frac{N^2-1}{12}$

The Hypergeometric random variable

- **Name**: Hypergeometric(M, N, n)
- **When to use**: When you are selecting a fixed number of items from a fixed size pool, containing a fixed number of desirable objects, without replacement and with order not mattering, and you are counting how many of the selected objects are desirable
- **Parameters**:
  - M: the number of desirable objects in the pool
  - N: the total number of objects in the pool
  - n: the number you are selecting ($n \leq M$)
- **Mass function**:

$$p_X(x) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} & \text{if } x = 0, 1, 2, 3, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

- **Statistics**:
  - $\mu = E(X) = n\frac{M}{N}$
  - $\sigma^2 = \mathrm{Var}(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{n-1}$

$\text{Hypergeometric}(M, N, n)$ is like $\text{Binomial}(n, M/N)$

- Viewing hypergeometric as "pick n, one after another", both are

$$X_1 + X_2 + \ldots + X_n$$

  with each $X_i \sim \text{Bernoulli}(M/N)$

- Difference
  - For $\text{Binomial}(n, M/N)$, X_i's are *independent*
  - For $\text{Hypergeometric}(M, N, n)$, they are *not*; $\Pr(X_n = 1)$ varies between

$$\frac{M}{N-(n-1)} \text{ and } \frac{M-(n-1)}{N-(n-1)}$$

    depending on previous choices

- If n small compared to N, M, not much difference here
- **Example**: When polling 1000 people (without replacement) out of 100,000,000 to see who they will vote for, can model situation with Binomial (easy) rather than Hypergeometric (harder)

The Poisson process

Events occur repeatedly over a period of time

- Occurrences in disjoint time intervals are independent
- Simultaneous occurrences are very rare
- The average number of occurrences per unit time is constant throughout the time period

The Poisson random variable

- **Name**: $\mathrm{Poisson}(\lambda)$
- **When to use**: When you are counting the number of occurrences of an event in unit time, when the occurrences satisfy the conditions of the Poisson process; or, when you are approximating $X \sim \mathrm{Binomial}(n, p)$ with n large, p small, np moderate
- **Parameter**:
  - λ: the average number of occurrences per unit time, or np
- **Mass function**:

$$p_X(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{if } x = 0, 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

- **Statistics**:
  - $\mu = E(X) = \lambda$
  - $\sigma^2 = \mathrm{Var}(X) = \lambda$