SPECIAL DISTRIBUTIONS

Some distributions are "special" because they are useful. They include: Poisson, exponential, Normal (Gaussian), Gamma, geometric, negative binomial, Binomial and hypergeometric distributions. We already saw and used exponential, Binomial and hypergeometric distributions. We will now explore the definitions and properties of the other "special" distributions.

Things to remember when learning probability distributions:

Mathematical Details	Application Context
1. Is it **continuous**, or **discrete**? What's the sample space?	1. Corresponding experiment? Cartoon example?
2. Density(mass) function? Parameter ranges?	2. Any common applications? Why is it "special" or useful?
3. Expected value formula? Variance?	3. Relationships to other distributions?

Bernoulli. Sample space $\{0, 1\}$, mean p, variance $p(1-p)$, and mass function

$$p_x = p^x(1-p)^{1-x}$$

Binomial. The number of successes in n Bernoulli(p) trials. Discrete random variable with sample space $\{0, \dots, n\}$, and mass function (with parameters n, p) given by

$$p_x = \binom{n}{k} p^x(1-p)^{n-x}$$

The mean is np and the variance is $np(1-p)$.

Multinomial (Generalized Binomial). Discrete random variable for the number of each of k types of outcomes in n trials. Sample space $\{0, ..., n\}^k$, and mass function (with parameters n,p_1,...,p_k where the $\sum p_i = 1$) given by

$$p_{x_1,\dots,x_k} = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

The marginals are binomial, thus the means are $E(X_i) = np_i$ and the variances are $Var(X_i) = np_i(1-p_i)$.

Hypergeometric. Discrete r.v. with sample space $\{0, ..., w\}$, and mass function (with parameters N, w, n) given by

$$p_x = \frac{\binom{w}{x}\binom{N-w}{n-x}}{\binom{N}{n}}$$

The mean is $n\,w/N$ and the variance is $n\,w/N(1-w/N)(N-n)/(N-1)$.

Generalized Hypergeometric. Discrete random variable with parameters n, n_1, ..., n_k, $\sum n_i = N$, with mass function

$$p_{x_1,\ldots,x_k} = \frac{\binom{n_1}{x_1}\cdots\binom{n_k}{x_k}}{\binom{N}{n}}$$

The marginals are Hypergeometric.

Uniform (Continuous). Discrete Continuous random variable with sample space {}, and density function

$$f_x = (b-a)^{-1}$$

The mean is $(b+a)/2$ and the variance is $(b-a)^2/12$.

Poisson distribution. Discrete random variable with $\lambda > 0$, mass function

$$P(X=k) = \frac{e^{-\lambda}(\lambda^k)}{k!} \text{ for } k = 0, 1, 2, \ldots.$$

The mean and variance are the same, namely $E(X) = Var(X) = \lambda$.

Poisson Approximation to Binomial distribution. Let $X \sim Bin(n,p)$ be a binomial random variable with number of trials n and probability of success p. Then, for large n and small p, that is when $n \to \infty$ and $p \to 0$ in such a way that $np = \lambda$ is held constant, we have

$$\lim_{n\to\infty, p\to 0} P(X=k) = \frac{e^{-np}(np)^k}{k!} = \frac{e^{-\lambda}(\lambda)^k}{k!}.$$

For large values of n, small p, we can therefore approximate the Binomial distribution with a Poisson distribution: $P(X=k) \approx \frac{e^{-np}(np)^k}{k!}$.

Poisson Model. Suppose events can occur in space or time in such a way that:

1. The probability that two events occur in the same *small* area or time interval is zero.
2. The events in disjoint areas or time intervals occur independently.
3. The probability than an event occurs in a given area or time interval T depends only on the size of the area or length of the time interval, and not on their location.

Poisson Process. Suppose that events satisfying the Poisson model occur at the rate λ per unit time. Let X(t) denote the number of events occuring in time interval of length t. Then

$$P(X=k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

X(t) is called *Poisson process* with rate λ.

Exponential Continuing from above, the waiting time Y between consecutive events has an exponential distribution with parameter λ (that is with mean $1/\lambda$), that is $P(Y > t) = e^{-\lambda t}, \;\; t > 0$, or equivalently,

$$f(t) = \lambda e^{-\lambda t}, \text{ for } t > 0.$$

The mean is $1/\lambda$ and the variance is $1/\lambda^2$.

Geometric and Negative Binomial distributions.

Geometric experiment: Toss a fair coin until the first H appears. Let X=number of tosses required for the first H. Then X has geometric distribution with probability of success 0.5.

Definition: Geometric distribution. A random variable X has a geometric distribution with parameter p if its pmf is

$$P(X=k)=(1-p)^{k-1}p, \text{ for } k=1,2,3,\ldots.$$

It is denoted $X \sim Geo(p)$. The mean and variance of a geometric distribution are $EX=1/p$ and $Var(X)=\frac{1-p}{p^2}$, respectively. The mgf of X is $M_X(t)=\frac{pe^t}{1-(1-p)e^t}$.

Memoryless property of geometric distribution. Let $X \sim Geo(p)$, then for any n and k, we have

$$P(X=n+k \mid X>n)=P(X=k).$$

Negative Binomial experiment. Think of geometric experiment performed until we get r successes. Let X = number of trials until we have r successes.

Definition: Geometric distribution. A random variable X has a negative binomial distribution with parameters r and p if its pmf is

$$p_X(k)=P(X=k)=\begin{pmatrix} k-1 \\ r-1 \end{pmatrix}(1-p)^{k-r}p^r, \text{ for } k=r,r+1,r+2,\ldots. \tag{1}$$

A common notation for X is $X \sim NegBin(r,p)$. The mean and variance of X are $EX=r/p$ and $VarX=\frac{r(1-p)}{p^2}$. The mgf of X is $M_X(t)=\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$.

Connection between negative binomial and geometric distributions. If $X_1,X_2,\ldots,X_r$ are iid $Geo(p)$, then $\sum_{i=1}^r X_i \sim NegBin(r,p)$.

The Gamma distribution.

Definition. The Gamma function. For any positive real number $r > 0$, the *gamma function* of r is denoted $\Gamma(r)$ and equal to

$$\Gamma(r) = \int_0^\infty y^{r-1} e^{-y} dy.$$

Theorem. Properties of Gamma function. The Gamma(r) function satifies the following properties:

1. $\Gamma(1) = 1$.
2. $\Gamma(r) = (r-1)\Gamma(r-1)$.
3. For r integer, we have $\Gamma(r) = (r-1)!$.

Definition of the $\Gamma(r, \lambda)$ random variable. For any real positive numbers $r > 0$ and $\lambda > 0$, a random variable with pdf

$$f_X(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \quad x > 0,$$

is said to have a Gamma distribution with parameters r and λ, denoted $X \sim \Gamma(r, \lambda)$.

Theorem: moments and mgf of a gamma distribution. If $X \sim \Gamma(r, \lambda)$ then

1. EX= r/λ.
2. Var(X)= r/λ^2.
3. Mgf of X is $M_X(t) = (1 - t/\lambda)^r$.

Theorem. Let $X_1, X_2, \ldots, X_n$ be iid exponential random variables with parameter λ, that is with mean $1/\lambda$. The the sum of X_i's has a gamma distribution with parameters n and λ. More precisely, $\sum_{i=1}^n X_i \sim \Gamma(r, \lambda)$.

Theorem. A sum of independent gamma random variables $X \sim \Gamma(r, \lambda)$ and $Y \sim \Gamma(s, \lambda)$ with the same λ has a gamma distribution with $r' = r + s$ and the same λ. That is $X + Y \sim \Gamma(r + s, \lambda)$.

Note: In a sequence of Poisson events occurring with rate λ per unit time/area, the waiting time for the r'th event has a $\Gamma(r, \lambda)$ distribution.