We discuss few probability distribution functions particularly relevant to astronomical research.

- Some are discrete with countable jumps while others are continuous.
- The mathematical properties of these distributions have been extensively studied, sometimes for over two centuries. These include moments, methods for parameter estimation, order statistics, interval and extrema estimation, two-sample tests, and computational algorithms.
- We pay particular attention to the Gaussian, Chi-square, Poisson and Pareto (or power law) distributions which are particularly important in astronomy.

The p.m.f.'s and p.d.f.'s have normalizations that may be unfamiliar to astronomers because their sum or total integral needs to be equal to unity.

For example, the astronomer's familiar power law distribution \( f(x) = bx^{-\alpha} \) is not a p.d.f. for arbitrary choices of \( \alpha \) and \( b \).

The Pareto distribution
\[
f(x) = \begin{cases} \alpha b^{\alpha} / x^{\alpha+1} & \text{for } x \geq b > 0 \\ 0 & \text{for } x < b \end{cases}
\]
has the correct normalization for a p.d.f.

Let \( X \) be a binomial \((n, p)\) random variable.

- If there are \( k \) successes, there must also be \( n - k \) failures.
- By independence, any single outcome with \( k \) success and \( n - k \) failures has probability \( p^k (1 - p)^{n-k} \).
- There are \( \binom{n}{k} \) such outcomes.

Therefore, the p.m.f. is
\[
f_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}.
\]

The mean and variance of a binomial \((n, p)\) random variable are \( np \) and \( np(1 - p) \), respectively.

In the special case of \( n = 1 \), we have a Bernoulli random variable.

\[
P \left( \frac{X - np}{\sqrt{np(1 - p)}} \leq x \right) \approx \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy
\]

A rule of thumb for normal approximation is \( np(1 - p) > 9 \).
### Multinomial distribution

- The multinomial distribution is a $k$-dimensional generalization of the binomial that provides $k > 2$ possible outcomes for each trial.
- Many applications to astronomy are evident: Type Ia, Ib, Ic, and II supernovae; Class 0, I, II and III pre-main sequence stars; Flora, Themis, Vesta and other asteroid families; etc.
- A random vector $\mathbf{N} = (N_1, \ldots, N_k)$ follows multinomial distribution with parameters $p_i > 0$ where $\sum_{i=1}^{k} p_i = 1$ and the number of trials occurring in each class is $N_i$ where $\sum_{i=1}^{k} N_i = n$ (total number of trials).
- If $\sum_{i=1}^{k} m_i = n$ and zero otherwise.

\[
P(N_1 = m_1, \ldots, N_k = m_k) = \frac{n!}{m_1! \ldots m_k!} p_1^{m_1} \ldots p_k^{m_k},
\]

### Poisson distribution

**Events** follow approximately a Poisson distribution if they are produced by Bernouilli trials, where the probability $p$ of occurrence is very small, the number $n$ of trials is very large, but the product $\lambda = np$ approaches a constant. This is quite natural for a variety of astronomical situations:

- A distant quasar may emit $10^{64}$ photons s$^{-1}$ in the X-ray band but the photon arrival rate at a telescope may only be $\sim 10^{-3}$ photons s$^{-1}$. The signal from a typical $10^4$ s observation may thus contain only $10^1$ of the $n \sim 10^{58}$ photons emitted by the quasar during the observation period, giving $p \sim 10^{-67}$ and $\lambda \sim 10^1$.

- A 1 cm$^2$ detector is subject to bombardment of cosmic rays, producing an astronomically uninteresting background. As the detector geometrically subtends $p \sim 10^{-18}$ of the Earth's surface area, and that these cosmic rays should arrive in a random fashion in time and across the detector, the Poisson distribution can be used to characterize and statistically remove this instrumental background.

### Multinomial distribution (contd.)

- The estimators for the class probabilities $p_i$ are the same as with the binomial parameter, $N_i/n$, with variances $np_i(1 - p_i)$.
- For example, an optical photometric survey may obtain a sample of 43 supernovae consisting of 16 Type 1a, 3 Type 1b and 24 Type II supernovae. The sample estimator of the Type Ia fraction is then $\hat{p}_1 = 16/43$.
- By using the multivariate Central Limit Theorem it can be proved that

\[
\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}
\]

has approximately a chi-square distribution with $k - 1$ d.f.
- The $\chi^2$ quantity is the sum of the ratio $(O_i - E_i)^2 / E_i$ where $O_i$ is the observed frequency of the $i$-th class and $E_i$ is the expected frequency given its probability $p_i$.
- The accuracy of the $\chi^2$ approximation can be poor for large $k$ or if any $p_i$ is small.

### Poisson distribution

If $X$ is a Poisson ($\lambda$) random variable where $\lambda > 0$, its p.m.f. is exactly

\[
f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x \in \{0, 1, 2, \ldots\}
\]

By the way, $\lambda$ does not have to be an integer, but $X$ does.

A Poisson ($\lambda$) random variable has mean $\lambda$ and variance $\lambda$.

Computing the Poisson distribution function:

\[
\frac{P(X = i + 1)}{P(X = i)} = \frac{\lambda}{i + 1}
\]

Start with $P(X = 0) = e^{-\lambda}$ .

Exponential distribution

- $X$ is an exponential ($\lambda$) r.v. if the PDF of $X$ is
  
  $$f(x) = \begin{cases} 
  \lambda e^{-\lambda x} & \text{if } x > 0 \\
  0 & \text{if } x \leq 0,
  \end{cases}$$

  where the parameter $\lambda > 0$.
  
  - $E(X) = \frac{1}{\lambda}$, $\text{Var}(X) = \frac{1}{\lambda^2}$.

- The exponential r.v. with mean $\theta$ has density and c.d.f.
  
  $$f(x) = \begin{cases} 
  \frac{1}{\theta} e^{-x/\theta} & \text{if } x > 0 \\
  0 & \text{if } x \leq 0,
  \end{cases} \quad \text{and} \quad F(x) = \begin{cases} 
  1 - e^{-x/\theta} & \text{if } x > 0 \\
  0 & \text{if } x \leq 0.
  \end{cases}$$

  [Link to Wikipedia page on Exponential distribution]

Erlang distribution

- $X$ is Erlang ($n, \lambda$) r.v. if the PDF of $X$ is
  
  $$f(x) = \begin{cases} 
  \frac{1}{(n-1)!} \lambda^n x^{n-1} e^{-\lambda x} & \text{if } x > 0 \\
  0 & \text{if } x \leq 0.
  \end{cases}$$

  Agner Krarup Erlang, a Danish mathematician and engineer, developed it to examine the number of telephone calls which might be made at the same time to the operators of the switching stations.

  Erlang is a special case of Gamma distribution.

  [Link to Wikipedia page on Erlang distribution]

Gamma distribution

- $X$ is Gamma ($\alpha, \lambda$) r.v if the PDF of $X$ is
  
  $$f(x) = \begin{cases} 
  \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} & \text{if } x > 0 \\
  0 & \text{if } x \leq 0.
  \end{cases}$$

  $$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx$$

  [Link to Wikipedia page on Gamma distribution]

Chi-square distribution with $k$ degrees of freedom is Gamma with $\alpha = k/2$, $\lambda = 1$.

  [Link to Wikipedia page on Chi-square distribution]

Poisson Process

- The most common origin of the Poisson distribution in astronomy is a counting process, a subset of stochastic point processes which can be produced by dropping points randomly along a line.

- The number of points till time $t$ follows Poisson distribution with intensity $\lambda t$.

- The distance between any two points is exponential with parameter $\lambda$.

- Successive distances are independent

- Total distance to the $n$-th point has Gamma distribution with parameters $(n, \lambda)$.

  [Link to Wikipedia page on Poisson distribution]
Normal distribution

The most celebrated distribution in probability is the standard normal.

- If $Z$ is standard normal, then $f_Z(z) = (1/\sqrt{2\pi})e^{-z^2/2}$.
- The c.d.f. even has its own symbol: $\Phi(z) = F_Z(z)$.
- In R, $\Phi(z)$ and $\Phi^{-1}(p)$ are `pnorm(z)` and `qnorm(p)`.

Let $Z$ be standard normal.
Then $\sigma Z + \mu$ is also normal for real numbers $\sigma$ and $\mu$.
What is the density of $\sigma Z + \mu$?
The result is a normal $(\mu, \sigma^2)$ random variable.
(We usually take $\sigma > 0$.)

http://en.wikipedia.org/wiki/Normal_distribution

Power law

Also known as Pareto distribution.

$$P(X > x) \propto \left( \frac{x}{x_{min}} \right)^{-\alpha}, \quad \alpha > 0$$

The Pareto probability distribution function (p.d.f.) can be qualitatively expressed as

$$P(x) = \frac{\text{shape}}{\text{location}} \left( \frac{\text{location}}{x} \right)^{\text{shape} + 1}.$$
Power law in astronomy

Power law appear in:
- the sizes of lunar impact craters
- intensities of solar flares
- energies of cosmic ray protons
- energies of synchrotron radio lobe electrons
- the masses of higher-mass stars
- luminosities of lower-mass galaxies
- brightness of extragalactic X-ray and radio sources
- brightness decays of gamma-ray burst afterglows
- turbulent structure of interstellar clouds
- sizes of interstellar dust particles

Some of these distributions are named after the authors of seminal studies:
- de Vaucouleurs galaxy profile
- Salpeter stellar Initial Mass Function (IMF)
- Schechter galaxy luminosity function
- MRN dust size distribution
- Larson’s relations of molecular cloud kinematics

The power law behavior is often limited to a range of values:
- the spectrum of cosmic rays breaks around $10^{15}$ eV and again at higher energies
- the Schechter function drops exponentially above a characteristic galaxy absolute magnitude $M^* \sim -20.5$
- the Salpeter IMF becomes approximately lognormal below $\sim 0.5 M_\odot$

See the notes for extensions of power law and multivariate Pareto.