Inference, Models and Simulation for Complex Systems — CSCI 7000-003
Lecture 1 — 26 August 2010
Prof. Aaron Clauset

1 Probability distributions

A probability density function (pdf) is defined as any function that satisfies the equation

$$1 = C \int_{-\infty}^{\infty} p(x) \mathrm{d}x \ . \tag{1}$$

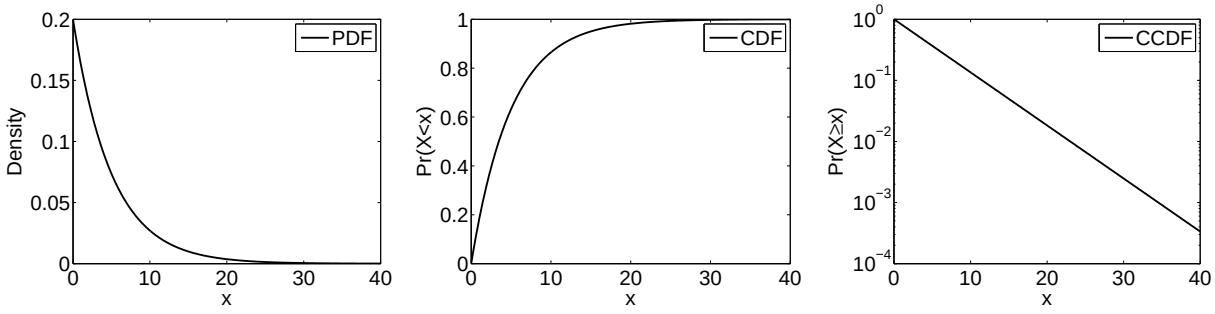
In many cases, the range of the function is somewhat less than $(-\infty, \infty)$. Generally, to convert a function into a pdf, we simply need to normalize it. (NB: not all functions can be normalized.) This means identifying the constant C that makes Eq. (1) true. For instance, consider an exponential function defined on the interval $[0, \infty)$,

$$1 = C \int_{0}^{\infty} \mathrm{e}^{-\lambda x} \mathrm{d}x \ .$$

Solving this yields $C = \lambda$, and thus the pdf for an exponential distribution is $\Pr(x) = \lambda \mathrm{e}^{-\lambda x}$. The cumulative distribution function (cdf) is defined similarly:

$$\begin{aligned} \Pr(X < x) &= C \int_{-\infty}^{x} p(y) \mathrm{d}y \\ &= \lambda \int_{0}^{x} \mathrm{e}^{-\lambda y} \mathrm{d}y \\ &= 1 - \mathrm{e}^{-\lambda x} \ . \end{aligned}$$

A complementary cdf is defined as $1 - \mathrm{cdf} = 1 - \Pr(X < x) = \Pr(X \geq x)$. For example, these plots those the exponential distribution's pdf, cdf and ccdf.



2 The Poisson process

2.1 Introduction

Suppose we have a stochastic system in which an event of interest occurs with small and constant probability q, and where the occurrence of events are independent (that is, the events are iid).

Such a process is called a Poisson process or a homogeneous Lévy process, and is a kind of memoryless Markov process. It is also called "pure-birth" process, and is the simplest of the family of models called "birth-death" processes. The name Poisson comes from the French mathematician Siméon Denis Poisson (1781–1840).

Examples might include

- the number of hikers passing some particular trailhead in the foothills above Boulder,
- a "death" event, e.g., of a computer program, an organism or a social group,
- the arrival of an email to your inbox.

Of course, these are probably not well modeled as Poisson processes: hikers tend to appear at certain times of day; processes might interact which can lead to correlations in their deaths; and emails are generated by people and people tend to synchronize and coordinate their behavior in complicated ways.

However, there are many things we can calculate about such a simple process. We'll do two:
(1) the distribution of "lifetimes" or delays between individual events.
(2) the distribution of the number of events observed within a given time window, and

There are several ways to do this; we'll use the *master equation* approach to study the continuous-time case. Other approaches, including the discrete case (see Section 5), yield equivalent results.

2.2 The continuous case

To begin:
(1) Let λ be the *arrival rate* of events (events per unit time).
(2) Let $P_x(t)$ denote the probability of observing exactly x events during a time interval t.
(3) Let $P_x(t + \Delta t)$ denote the probability of observing x events in the time interval $t + \Delta t$.
(4) Thus, by assumption, $q = P_1(\Delta t) = \lambda \Delta t$ and $1 - q = P_0(\Delta t) = 1 - \lambda \Delta t$.

For general $x > 0$ and small Δt, this can written mathematically as

$$P_x(t + \Delta t) = P_x(t)P_0(\Delta t) + P_{x-1}(t)P_1(\Delta t) \tag{2}$$
$$= P_x(t)(1 - \lambda \Delta t) + P_{x-1}(t)\lambda \Delta t \ . \tag{3}$$

In words, either we observe x events over t and no events over the Δt *or* we observe $x - 1$ events over t and exactly one event over the Δt.

With a little algebra, this can be turned into a difference equation

$$\frac{P_x(t + \Delta t) - P_x(t)}{\Delta t} = \lambda P_{x-1}(t) - \lambda P_x(t) \ ,$$

and letting $\Delta t \to 0$ turns it into a differential equation

$$\frac{\mathrm{d}P_x(t)}{\mathrm{d}t} = \lambda P_{x-1}(t) - \lambda P_x(t) \ . \tag{4}$$

When $x = 0$, i.e., there are no events over $t + \Delta t$, the first term of Eq. (4) can be dropped:

$$\frac{\mathrm{d}P_0(t)}{\mathrm{d}t} = -\lambda P_0(t) \ .$$

This is a simple ordinary differential equation (ODE) and admits a solution $P_0(t) = C\mathrm{e}^{-\lambda t}$, where it can be shown that $C = 1$ because $P_0(0) = 1$.

Importantly, $P_0(t)$ is the distribution of waiting times between events, because it's the distribution of times during which no events occur. When an event represents the "death" of an object, this is the distribution of object lifetimes and is our first result.

To get our second result, take Eq. (4), set $x = 1$ and substitute in our expression for $P_0(t)$.

$$\frac{\mathrm{d}P_1(t)}{\mathrm{d}t} = \lambda \mathrm{e}^{-\lambda t} - \lambda P_1(t) \ .$$

This gives a differential equation for $P_1(t)$. Solving this (another simple ODE) yields a solution $P_1(t) = \lambda t \mathrm{e}^{-\lambda t}$ (with boundary condition $P_0(0) = 0$). For general x, it yields

$$P_x(t) = \frac{(\lambda t)^x}{x!} \mathrm{e}^{-\lambda t} \ . \tag{5}$$

Finally, to get the probability of observing exactly x events per unit time, take $t = 1$ in Eq. (5). This yields

$$P_x = \frac{\lambda^x}{x!} \mathrm{e}^{-\lambda} \ , \tag{6}$$

which is the Poisson distribution, our second result.

2.3 The discrete case

We won't solve the entire discrete case, but we'll point out a few important connections. First, for a finite number of trials n where the probability of an event is q, the distribution of the number of events x is given by the binomial distribution

$$\Pr(X = x) = \binom{n}{x} q^x (1-q)^{n-x} \tag{7}$$

When q is very small, the binomial distribution is approximately equal to the Poisson distribution.

$$\Pr(X = x) = \frac{(qn)^x}{x!} \mathrm{e}^{-(qn)} \ , \tag{8}$$

where $\lambda = qn$.

The distribution of lifetimes follows the discrete analog of the exponential distribution, which is called the geometric distribution

$$\begin{aligned} \Pr(X = x) &= 1 - (1-q)^x \\ &\approx \mathrm{e}^{-qx} \end{aligned} \tag{9}$$

3 The exponential distribution

Suppose now that we observe some empirical data on some object lifetimes. If we assume that the data were generated by a Poisson-type process, how can we infer the underlying parameter λ directly from the observed data?

To do this, we'll introduce a technique called *maximum likelihood*, which was popularized by R. A. Fisher in the early 1900s, but actually first used by notables like Gauss and Laplace in the 18th and 19th centuries.

Recall that the (continuous) exponential distribution has the form

$$\Pr(x) = \lambda \mathrm{e}^{-\lambda x} \ , \tag{10}$$

and let $\{x_i\} = \{x_1, x_2, \ldots, x_n\}$ denote our observed lifetime data. The likelihood of these data under the exponential model is defined as

$$\begin{aligned} \mathcal{L}(\{x_i\} \,|\, \vec{\theta}) &= \prod_{i=1}^{n} \Pr(x_i) \\ \mathcal{L}(\{x_i\} \,|\, \lambda) &= \prod_{i=1}^{n} \lambda \mathrm{e}^{-\lambda x_i} \ , \end{aligned}$$

where we substitute the particular model parameter λ for the generalized parameter $\vec{\theta}$ once we substitute the particular probability distribution for the model we're studying. (NB: This step is entirely general and only requires assuming that your data are iid.)

Our goal now is to find the value of λ, denoted $\hat{\lambda}$, that *maximizes* this expression. Equivalently, we can find the value that maximizes the logarithm of the expression. (This works because the log is a monotonic function, and thus doesn't move the location of the maximum.) Thus,

$$\begin{aligned} \ln \mathcal{L}(\{x_i\} \,|\, \lambda) &= \ln \prod_{i=1}^{n} \lambda \mathrm{e}^{-\lambda x_i} \\ &= \sum_{i=1}^{n} \ln\Big(\lambda \mathrm{e}^{-\lambda x_i}\Big) \\ &= \sum_{i=1}^{n} \Big(\ln \lambda + \ln \mathrm{e}^{-\lambda x_i}\Big) \end{aligned}$$

$$= \sum_{i=1}^{n} (\ln \lambda - \lambda x_i)$$

$$\ln \mathcal{L}(\{x_i\} \mid \lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i \ . \tag{11}$$

Eq. (11) is called the *log-likelihood function* and is useful for a wide variety of tasks. It appears in Bayesian statistics, frequentist statistics, machine learning methods, etc. Often, it can be written down exactly, as in this case, but sometimes, we can only write down a function that is proportional to the log-likelihood function (see Markov chain Monte Carlo [MCMC] algorithms).

Singe Eq. (11) is so simple, we can find the location of the maximum by taking the derivative with respect to λ, setting the resulting expression equal to zero, and then solving for λ. When the expression is not simple, numerical methods can find the location of the maximum (see the Nelder-Mead method, also called the "simplex" method, among many other techniques).

$$0 = \frac{\partial}{\partial \lambda} \ln \mathcal{L}(\{x_i\} \mid \lambda)$$

$$0 = \frac{\partial}{\partial \lambda} \left(n \ln \lambda - \lambda \sum_{i=1}^{n} x_i \right)$$

$$0 = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

$$0 = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

$$\hat{\lambda} = 1 \Bigg/ \frac{1}{n} \sum_{i=1}^{n} x_i = 1 \, / \langle x_i \rangle \quad . \tag{12}$$

Eq. (12) is called the *maximum likelihood estimator* (MLE) and can be shown to have many nice properties; one is *asymptotic consistency*, in which as $n \to \infty$, $\hat{\theta} \to \theta$ almost surely. We revisit this particular property in the problem set.

Using the likelihood function, we can also derive an estimate of the uncertainty or standard error in our parameter estimate, so that when we report our parameter estimate using real data, we say $\hat{\theta} \pm \hat{\sigma}$. Generally, the variance $\hat{\sigma}^2 = 1/I(\theta)$ where $\partial^2 \mathcal{L}(\hat{\theta}) / \partial \theta^2 \to I(\theta)$, and $I(\theta)$ is the Fisher information at θ. (The Fisher Information basically captures the width of the curvature of the likelihood function at the maximum; the more narrow the function, the more certain our estimate.) For the exponential distribution, $\hat{\sigma} = \hat{\lambda}/\sqrt{n}$.

4 Simulations

A Poisson process is easy to simulate numerically, especially in the discrete case. Here's some Matlab code that does this and generates the results shown in Figure 1.

```
n = 10^3;  q = 5/n;  lambda = q*n;
r=(1:20)';

x = zeros(length(r),1);     % analytic Poisson distribution
x(1) = exp(-lambda)*lambda; % constructed via tail-recursion
for i=2:length(r)           %
    x(i) = x(i-1)*lambda/i; %
end;

M = rand(n,n)<q;            % n trials, each with n coin tosses
y = sum(M);                 % compute counts of events per trial
h = hist(y,(1:20))./n;      % convert counts into a histogram

figure(1);
g=bar((1:20),h); hold on;
plot(r,x,'ro','MarkerFaceColor',[1 0 0],'MarkerSize',8);  hold off;
set(g,'BarWidth',1.0,'FaceColor','none','LineWidth',2);
set(gca,'FontSize',16,'XLim',[1/2 17],'XTick',(1:2:20),'YLim',[0 0.22]);
ylabel('Proportion','FontSize',16);
xlabel('Number','FontSize',16);
k=legend('\lambda=5, n=1000','Expected'); set(k,'FontSize',16);

z = zeros(n,1);             % tabulate time-to-first event
for i=1:n                   % for each trial
    if sum(M(:,i))>0, z(i) = find(M(:,i)==1,1,'first'); end;
end;
z(z==0) = [];               % clear out instances where nothing happened

figure(2);
semilogy(sort(z),(length(z):-1:1)./length(z),'k-','LineWidth',2); hold on;
semilogy(sort(z),exp(-q*sort(z)),'r--','LineWidth',2); hold off;
set(gca,'FontSize',16);
xlabel('Waiting time, t','FontSize',16);
ylabel('Pr(T\geqt)','FontSize',16);
k=legend('\lambda=5, n=1000','Expected'); set(k,'FontSize',16);
```

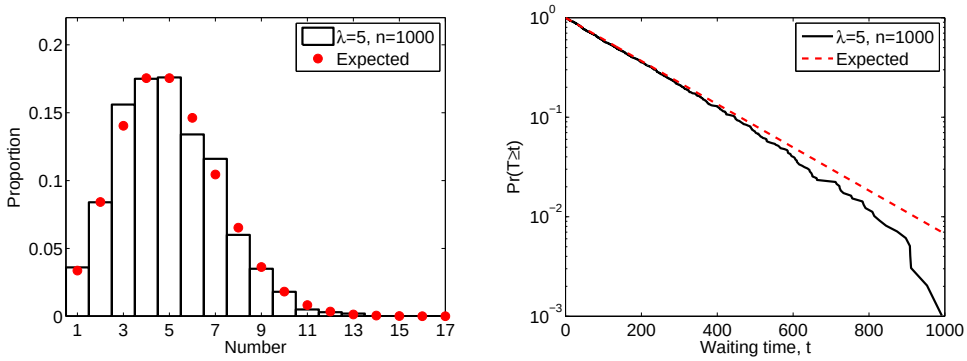
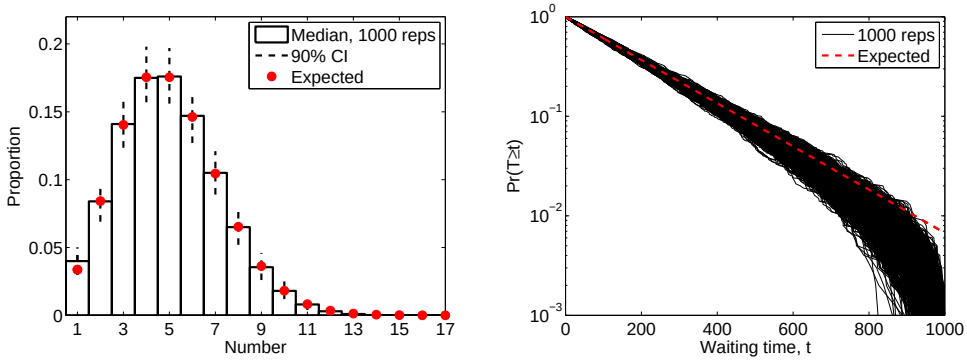


Figure 1: (A) The distribution for $n = 1000$ trials of a Poisson process with $\lambda = 5$, along with the expected counts for such a process, from Eq. (6). (B) The waiting-time distribution for the delay until the first event, for the same trials, along with the expected distribution.

If we apply our MLE to these data, we find $\hat{\lambda} = 0.0051$, which is very close to the true value of $q = 0.0050$. (NB: I'm abusing my notation a little here, by mixing λ and q.)

Notice that the observed counts (Fig. 1a) tend to deviate a little from the expected counts. Since the counts are themselves random variables, this is entirely reasonable. But, how much deviation should we expect to observe when we observe data drawn from a Poisson process?

Repeating the simulation m times, we can estimate a distribution for each count and put error bars on the expected values. Figure 4a shows the results for the counts, and Fig. 4b shows the variation in the distribution of waiting times. Note that this distribution bends downward close to $t = 1000$. This is because of a finite-size effect imposed by flipping only 1000 coins for each trial.



5 Alternative derivation of Poisson distribution

Consider a process in which we flip a biased coin, where the probability that the coin comes up 1 is q (an event occurs) and the probability of 0 is $(1-q)$ (an event does not occur). From the binomial theorem, we know that the distribution of the number of events (the number of 1s) in a long sequence of coin flips follows the binomial distribution

$$\Pr(X = x) = \binom{n}{x} q^x (1-q)^{n-x} \ , \tag{13}$$

where x is the number of events and n is the number of trials (and technically $0 \leq x \leq n$).
Recall that, by assumption, q is small. In this limit, we can simplify the binomial distribution in the following way.

To begin, rewrite $q = \lambda/n$ where λ is the expected number of events in n trials and recall from combinatorics that $\binom{n}{x} = \frac{n!}{(n-x)!\,x!}$:

$$\lim_{n\to\infty} \Pr(X = x) = \lim_{n\to\infty} \binom{n}{x} q^x (1-q)^{n-x} \tag{14}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-x)!\,x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \tag{15}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-x)!\,x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n} \left(1 - \frac{\lambda}{n}\right)^{-x} \ . \tag{16}$$

This form is convenient because we can use a basic equality from calculus

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = \mathrm{e}^{-\lambda} \ . \tag{17}$$

which allows us to simplify the second-to-last term in Eq. (16):

$$\lim_{n\to\infty} \Pr(X = x) = \lim_{n\to\infty} \frac{n!}{(n-x)!\,x!} \left(\frac{\lambda}{n}\right)^x \mathrm{e}^{-\lambda} \left(1 - \frac{\lambda}{n}\right)^{-x} \ . \tag{18}$$

Notice also that the last term is going to 1 because x is some constant, while $n \to \infty$. Thus, we can drop the last term, which yields

$$\lim_{n\to\infty} \Pr(X = x) = \lim_{n\to\infty} \frac{n!}{(n-x)!\,x!} \left(\frac{\lambda}{n}\right)^x \mathrm{e}^{-\lambda} \tag{19}$$

$$= \lim_{n\to\infty} \left(\frac{n!}{(n-x)!\,n^x}\right) \frac{\lambda^x}{x!} \mathrm{e}^{-\lambda} \tag{20}$$

Note that the left-hand term $\to 1$. This can be seen by observing that

$$\lim_{n\to\infty} \frac{n!}{(n-x)!\, n^x} = \lim_{n\to\infty} \frac{n}{n}\frac{(n-1)}{n}\frac{(n-2)}{n}\cdots\frac{(n-x+1)}{n} \tag{21}$$

$$= 1 \cdot 1 \cdot 1 \cdots 1 \tag{22}$$

for constant $x \geq 1$, where we apply the limit to each term individually (this is allowed because there are a finite number of terms). Thus, we have our main result, the Poisson distribution:

$$\Pr(X = x) = \frac{\lambda^x}{x!}\mathrm{e}^{-\lambda} \ . \tag{23}$$

An easy way to derive the distribution of waiting times between events for a Poisson process is to recall that λ is the expected number of events per *unit* time. Thus, if we rescale $\lambda \to \lambda t$, we have the number of events over some time span t. Setting $x = 0$ lets us consider waiting at least t time units see the first event. This yields

$$\begin{aligned}\Pr(X = 0, T > t) &= \frac{(\lambda t)^0}{0!}\mathrm{e}^{-\lambda t} \\ &= \frac{(\lambda t)^0}{0!}\mathrm{e}^{-\lambda t} \\ &= \mathrm{e}^{-\lambda t} \ .\end{aligned}$$

To get the distribution for waiting exactly t time units, we now simply differentiate with respect to time the expression $1 - \Pr(X = 0, T > t)$, which yields the exponential distribution $P(T = t) = \lambda\mathrm{e}^{-\lambda t}$.