ORF 245 Fundamentals of Statistics
Chapter 2
Random Variables

Robert Vanderbei

Fall 2014

Slides last edited on September 22, 2014

Definitions

Random Variable: A real-valued function on Ω is called a *random variable*.
We use capital letters such as X for random variables.
The notation $X \leq x$ is shorthand for the event $\{\omega \in \Omega \mid X(\omega) \leq x\}$.

Cumulative Distribution Function (cdf):

$$F(x) = P(X \leq x), \text{ for all } -\infty < x < \infty.$$

Probability Mass Function (aka Frequency Function): This only works for random variables that take on a discrete set of values, $x_1, x_2, \ldots$:

$$p(x_i) = P(X = x_i), \text{ for all } i = 1, 2, \ldots.$$

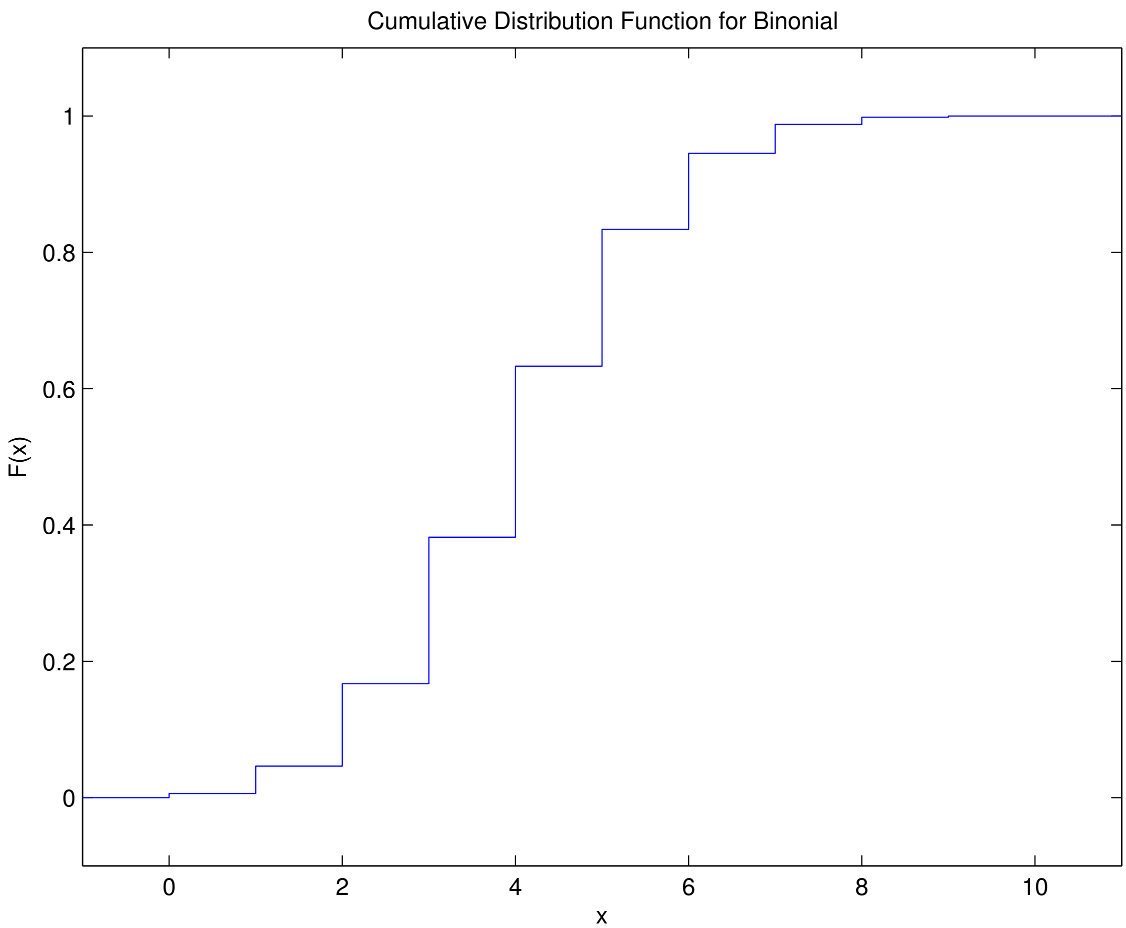
Independence: Two random variables, X and Y, are *independent* if every event expressible in terms of X alone is independent of every other event expressible in terms of Y alone. In particular,

$$P(X \leq x \text{ and } Y \leq y) = P(X \leq x)\, P(Y \leq y).$$

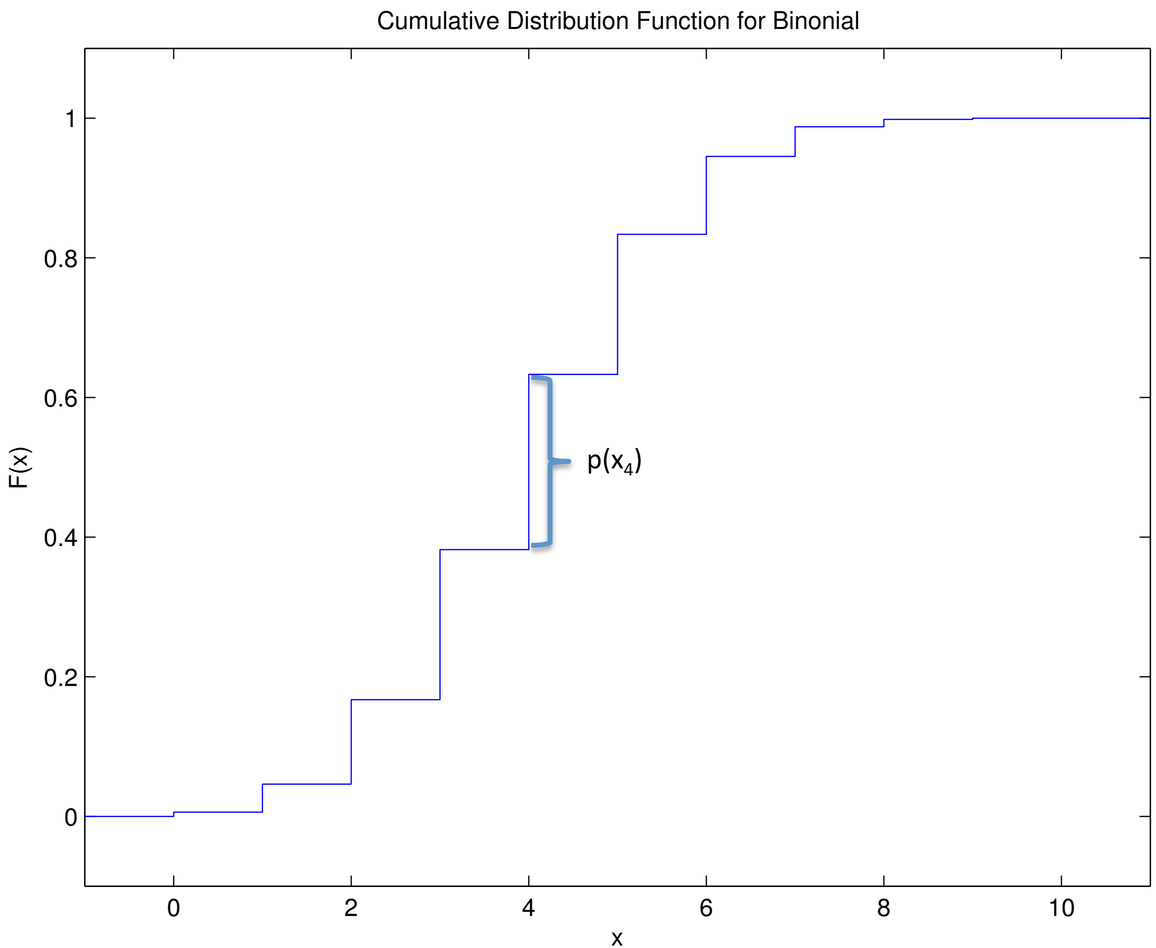
If the random variables are discrete, then

$$P(X = x_i \text{ and } Y = y_j) = P(X = x_i)\, P(Y = y_j).$$

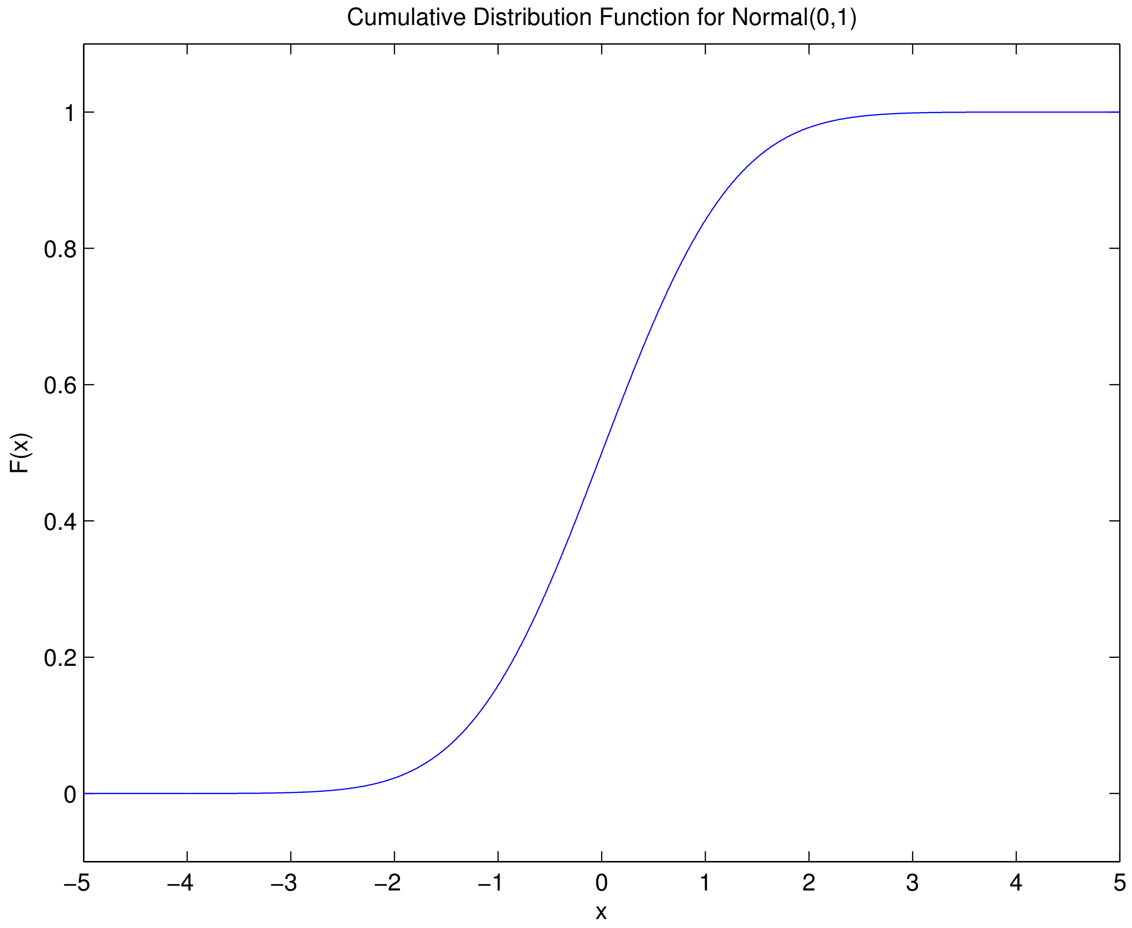
Example of a CDF – Discrete Case



Example of a CDF – Discrete Case



Example of a CDF – Continuous Case



Discrete Distributions

Bernoulli Distribution

A Bernoulli random variable X takes on just two values: zero or one.

At the risk of confusing notations, we usually let

$$p = P(X = 1)$$

and

$$q = P(X = 0) = 1 - p.$$

Bernoulli random variables are closely associated with "events". Let $A \subset \Omega$ be some event in some abstract sample space Ω. Let

$$X(\omega) = \begin{cases} 1 & \text{for } \omega \in A \\ 0 & \text{for } \omega \notin A \end{cases}$$

Such "indicator" random variables are often denoted by

$$X(\omega) = 1_A(\omega)$$

which means "one if event A happens, zero otherwise".

Bernoulli random variables often represent "success" vs. "failure" of an experiment.

Binomial Distribution

An experiment is performed n times.

Each time the experiment is "performed" is completely independent of each other time.

We assume that the experiment is a "success" with probability p and a "failure" with probability $q = 1 - p$ (i.e., each experiment is described by a Bernoulli random variable, say Y_j).

Let X denote the number of successes (i.e., $X = \sum_{j=1}^{n} Y_j$).

The random variable X has a binomial distribution:

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} p^k q^{n-k}.$$

Geometric Distribution

Again, we consider a sequence of independent Bernoulli trials.

In this case, there is no upper bound on how many trials will be performed.

Let X denote the number of trials that must be performed until a “success” occurs.

Such a random variable has a geometric distribution:

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

Sanity check: the probabilities should sum to one...

$$\sum_{k=1}^{\infty} p(k) = \sum_{k=1}^{\infty} pq^{k-1} = p \sum_{k=0}^{\infty} q^k = \frac{p}{1-q} = \frac{p}{p} = 1 \qquad \text{YES!}$$

Geometric series. Geometric random variable. A coincidence? No!

Negative Binomial Distribution

Same set up as before. But, now let X denote the number of trials required until the r-th success (where r is some given integer).

The event $\{X = k\}$ happens when in the first $k-1$ trials there were exactly $r-1$ successes and on the k-th trial there was also a success.

Hence,

$$p(k) = P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} = \binom{k-1}{r-1} p^r q^{k-r}, \qquad \text{for } k = r, r+1, \ldots.$$

Sanity check: the probabilities should sum to one...

$$\sum_{k=r}^{\infty} p(k) = \sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r q^{k-r} = \cdots == 1$$

Details left to reader!

Poisson Distribution

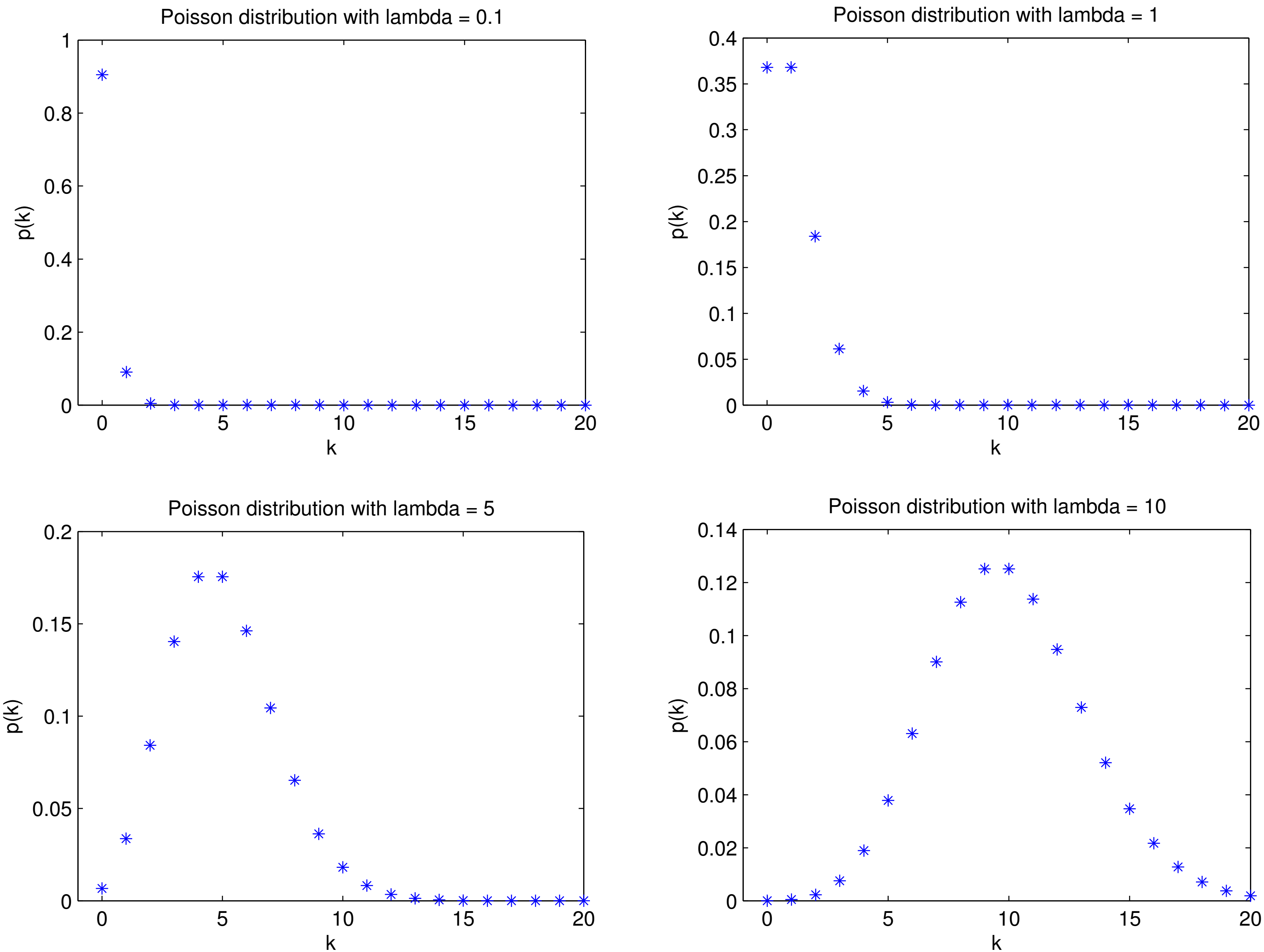
Start with a Binomial distribution with very large n and very small p.
Let $\lambda = pn$.
Let $n \to \infty$ and $p \to 0$ in such a way that λ remains constant.
The limiting distribution is called the Poisson Distribution:

$$\begin{aligned}
p(k) &= \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k} \\
&= \lim_{n\to\infty} \frac{1}{k!} \frac{n(n-1)\cdots(n-k+1)}{n^k} \lambda^k \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k} \\
&= \frac{\lambda^k}{k!} e^{-\lambda}
\end{aligned}$$

Sanity check: the probabilities should sum to one...

$$\sum_{k=0}^{\infty} p(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1 \qquad \text{Yes!}$$

It is fair to say that this Poisson distribution is the most important of all discrete distributions.


Poisson distribution with lambda = 0.1
Poisson distribution with lambda = 1
Poisson distribution with lambda = 5
Poisson distribution with lambda = 10
p(k)
k
0
5
10
15
20
0
0.2
0.4
0.6
0.8
1
0.05
0.1
0.15
0.25
0.3
0.35
0.02
0.04
0.06
0.08
0.12
0.14

Matlab Code

```
k = 0:20;
lambda = 10;
p = exp(-lambda) * lambda.^k ./ factorial(k) ;
figure(6);
plot(k,p,'b*');
xlim([-1 20]);
xlabel('k');
ylabel('p(k)');
title('Poisson distribution with lambda = 10');
```

Continuous Distributions

If $F(x) = P(X \leq x)$ is a continuous function of x, then the random variable X is said to have a *continuous distribution*.

Except in very special cases, a continuous increasing function is also differentiable.

Hence, we will assume that $F(x)$ has a derivative $f(x)$.

And, since $F(-\infty) = 0$, we can write

$$F(x) = \int_{-\infty}^{x} f(\xi) d\xi.$$

The function $f(x)$ is called the *density function*.

The area under the density function gives us probabilities:

$$P(a < X \leq b) = F(b) - F(a) = \int_{a}^{b} f(x) dx.$$

Note that

$$P(X = c) = \int_{c}^{c} f(x) dx = 0.$$

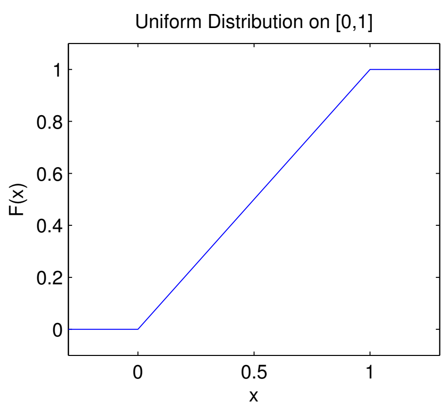
Hence,

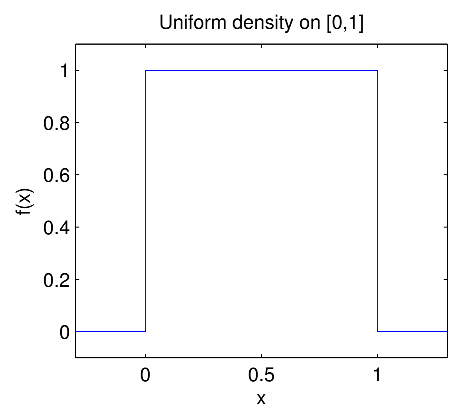
$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$$

Uniform Distribution

Pick a number "at random" from the interval $[0, 1]$.
The density function is

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & x < 0 \ \text{ or } \ x > 1 \end{cases}$$





If, instead of the interval $[0, 1]$, we pick a number at random from the interval $[a, b]$, then the density function is

$$f(x) = \begin{cases} 1/(b-a), & a \leq x \leq b \\ 0, & x < a \ \text{ or } \ x > b \end{cases}$$

Exponential Distribution

Exponential random variables describe random temporal events such as "how long until the next customer arrives?" We often use T instead of X for an exponential random variable.

The exponential density function is

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

The cumulative distribution function is easy to compute:

$$F(t) = \int_{-\infty}^{t} f(u)du = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

Memorylessness:

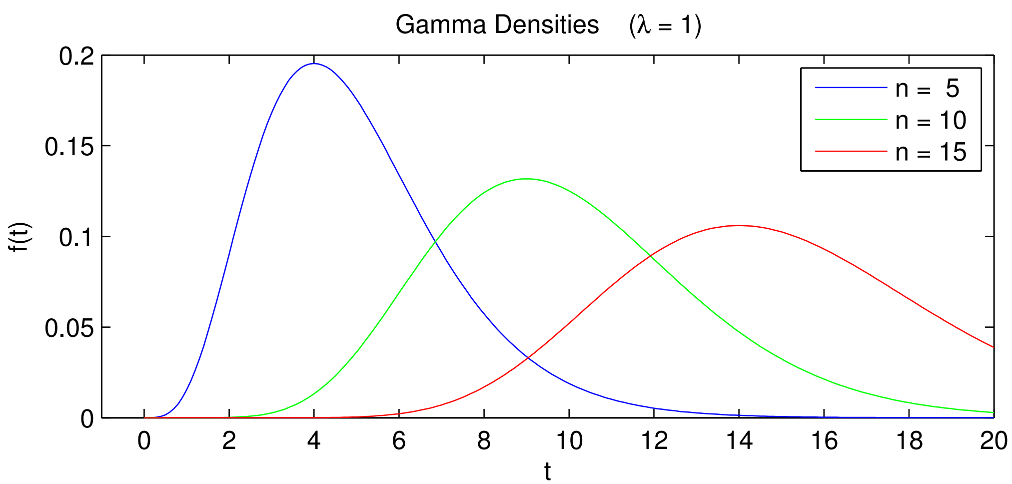
$$\begin{aligned} P(T > t + s \mid T > s) &= \frac{P(T > t + s \text{ and } T > s)}{P(T > s)} = \frac{P(T > t + s)}{P(T > s)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} \\ &= P(T > t) \end{aligned}$$

Gamma Distribution

A Gamma random variable is often used for a generic example of a *nonnegative* random variable. Its density function depends on two (positive) parameters, n and λ:

$$f(t) = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}, \qquad t \geq 0$$

Note: Gamma with $n = 1$ is the same as exponential.



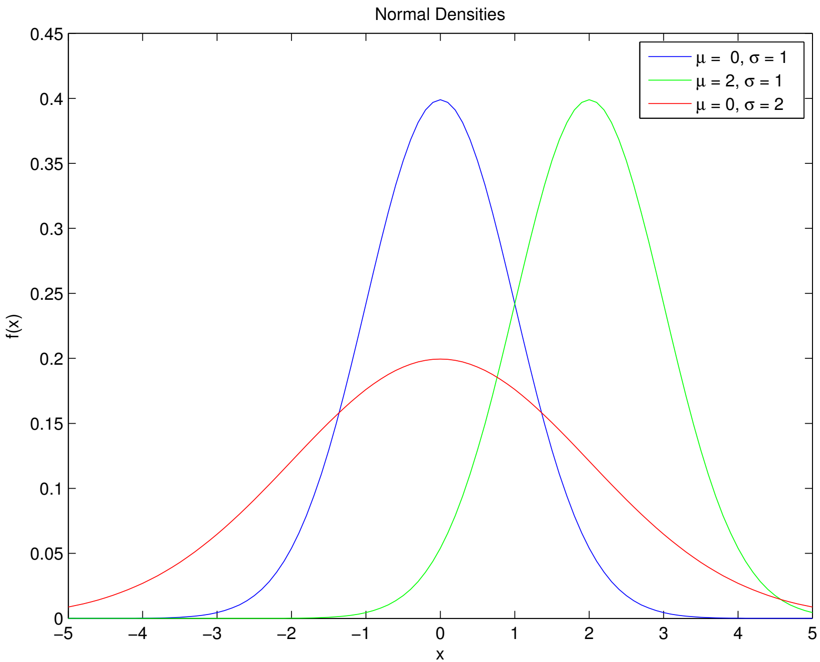
Normal (aka Gaussian) Distribution

A Normal random variable is often used as a generic symmetric random variable; i.e., the *bell-shaped curve*. Its density function depends on two parameters, μ and σ:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$$

Parameter μ is called the *mean* and parameter σ is called the *standard deviation*.

Note: the density peaks at $x = \mu$ and its "spread" increases with σ.



Matlab Code

```
dx = 0.1;
x = (-50:50)*dx;
mu = 2;
sigma = 1;
f = (1/(sqrt(2*pi)*sigma))*exp(-(x-mu).^2/(2*sigma^2));
sum(f)*dx
plot(x,f,'b-');
```

Functions of a Random Variable

Suppose that $X \sim N(\mu, \sigma^2)$.

What's the distribution of $Y = aX + b$?

Suppose (for convenience) that $a > 0$.

It's best to work with cdf's:

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right).$$

To find the density, we differentiate using the chain rule...

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X\left(\frac{y-b}{a}\right) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right) = \frac{1}{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi} a\sigma} e^{-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}} \end{aligned}$$

From this last expression, we see that $Y \sim N(a\mu + b, a^2\sigma^2)$.

Velocity/Energy

Suppose we live in a one-dimensional world (higher dimensions will come later) and that a certain particle has mass m and a random velocity $V \sim N(0, \sigma^2)$.

Find the distribution of its energy: $E = \frac{1}{2}mV^2$

First, compute the cdf:

$$\begin{aligned} F_E(x) &= P(E \le x) = P\left(mV^2/2 \le x\right) = P\left(-\sqrt{2x/m} \le V \le \sqrt{2x/m}\right) \\ &= F_V\left(\sqrt{\frac{2x}{m}}\right) - F_V\left(-\sqrt{\frac{2x}{m}}\right) \end{aligned}$$

Differentiating, we compute the density:

$$\begin{aligned} f_E(x) &= \sqrt{\frac{2}{m}}\frac{1}{2}x^{-1/2}\left(f_V\left(\sqrt{2x/m}\right) + f_V\left(-\sqrt{2x/m}\right)\right) = \sqrt{\frac{2}{m}}x^{-1/2}f_V\left(\sqrt{2x/m}\right) \\ &= \sqrt{\frac{2}{m}}x^{-1/2}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x}{m\sigma^2}} \qquad \Longleftarrow \text{Gamma w/ params } \alpha = 1/2,\ \lambda = 1/m\sigma^2 \end{aligned}$$

Simulation (Proposition 2.3.D)

Let U be a random variable that's uniform on $[0, 1]$.

Let $F(x)$ be a cumulative distribution function.

Because F is increasing, it has an inverse F^{-1}.

Let $X = F^{-1}(U)$.

Show that X is a random variable whose cdf if $F(x)$.

Compute:

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$$

If we want X to be exponential, then $F^{-1}(u) = -\ln(1-u)/\lambda$. We can use this to generate random exponential random variables from random uniformly distributed random variables.