MSL / FISH 604
Module 2 (part 3) : Basic statistical concepts

Instructor: Franz Mueter

Lena Point, Rm 315

796-5448

fmueter@alaska.edu

MSL
FISH 604

Review / Preview

- Probability and probability distributions
- Discrete distributions
  - Bernoulli
  - Binomial
  - Multinomial
  - Poisson
  - Negative binomial
- Next: continuous distributions
  - Uniform
  - Exponential
  - Normal or Gaussian
  - Log-normal
  - Gamma

Today

Objective and outcomes

Objectives

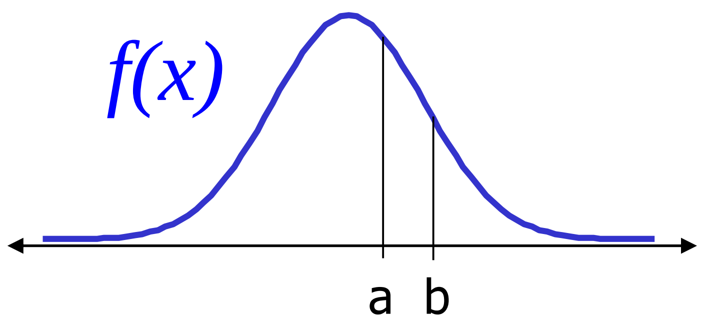
- Review / introduce <u>continuous</u> probability distributions (probability density functions or pdf)

Learning outcomes

- Know the major continuous probability distributions and their basic characteristics & uses:
  - Shape
  - Support (what values can the random variable take)
  - Parameters and their interpretation
  - Uses in science/ecology
- Understand how these distributions are used in statistics and be able to apply them

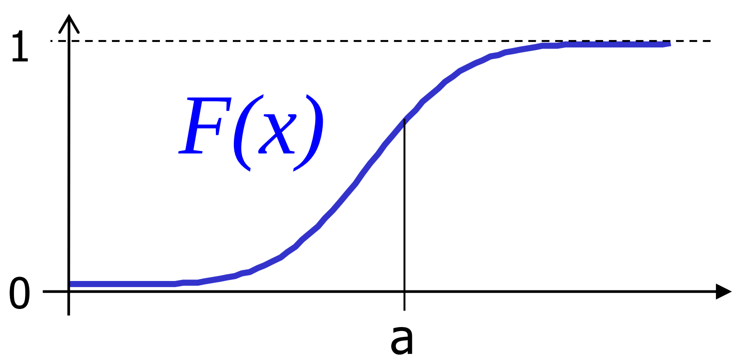
Continuous distributions

<u>Probability density function</u>: pdf or *f(x)*:



$$P(a < X < b) = \int_a^b f(x)dx$$

<u>Cumulative distribution function</u>: cdf of *F(x)*



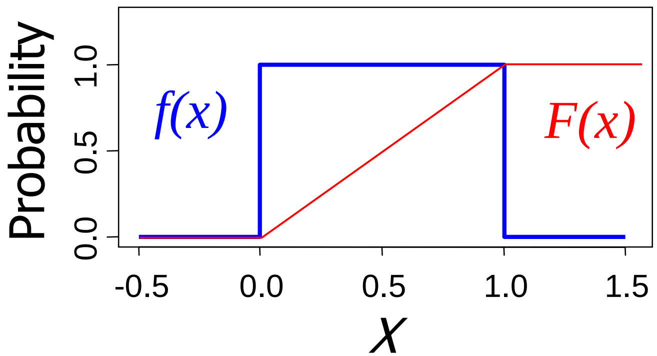
$$F(a) = P(X < a) = \int_{-\infty}^{a} f(x)dx$$

Uniform

- Uniform density on Interval [a,b]:

U(a,b)

U(0,1)



- Mean:

- Variance:

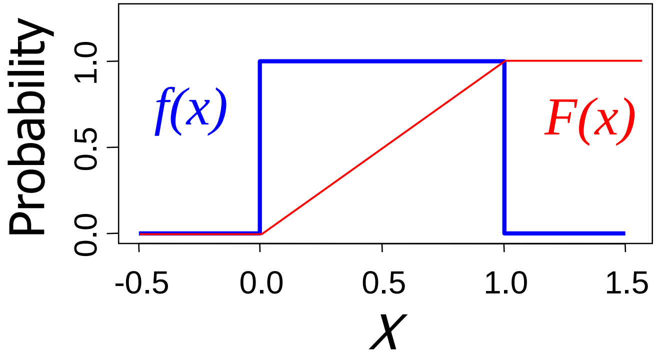
Uniform

- Uniform density on Interval [a,b]:

U(a,b)

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

U(0,1)



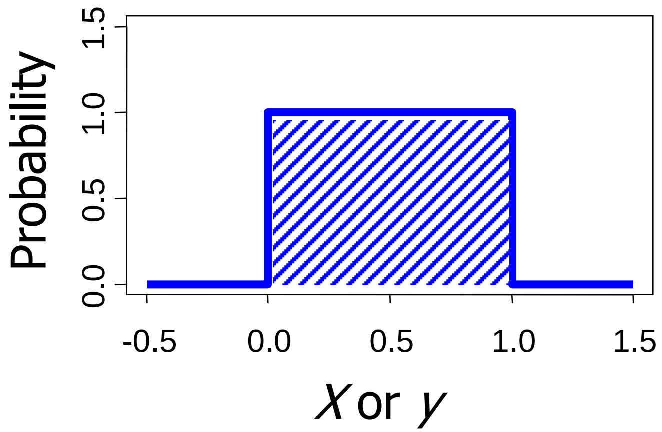
- Mean: $$E(x) = (a+b)/2$$

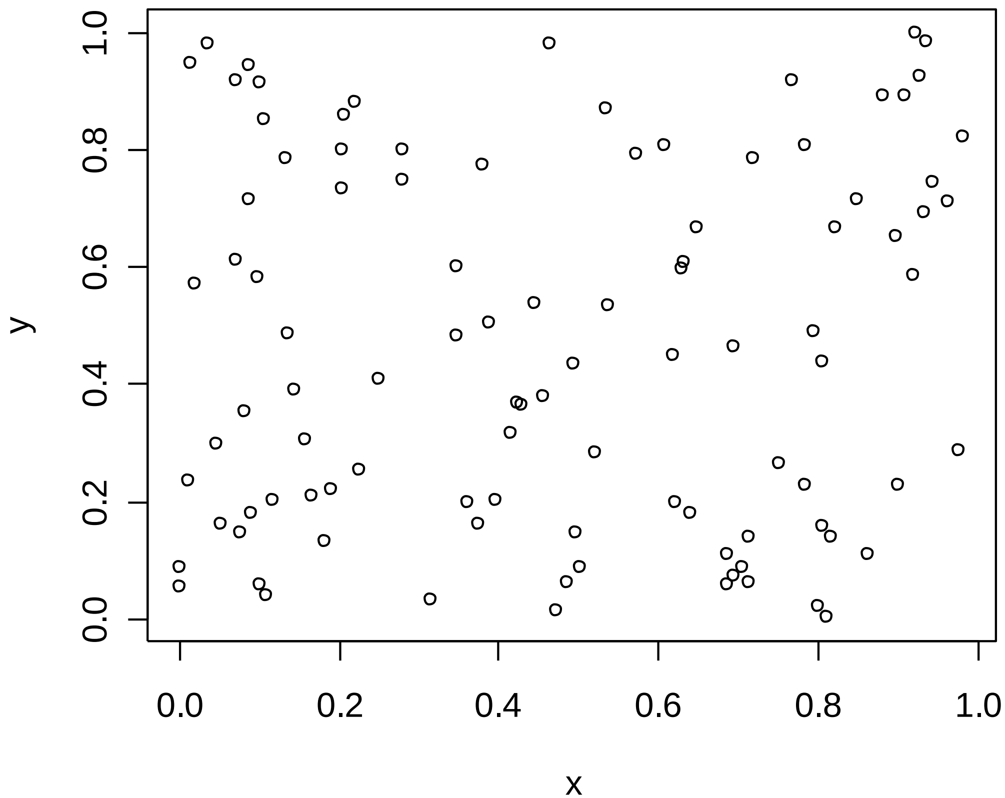
- Variance: $$Var(x) = (b-a)^2/12$$

Example: Uniform

Spatial randomness:

- $x \sim U(0,1)$, $y \sim U(0,1)$

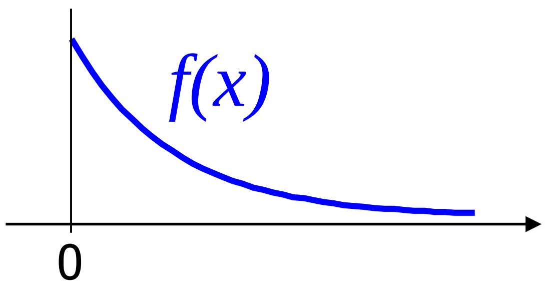




Exponential

- Often modeled to use life times or waiting times between events
- Completely specified by a single parameter λ (or *t*)
- Probability density function:

$$f(x) = \lambda e^{-\lambda x}, \qquad x \geq 0$$



- Mean and variance:

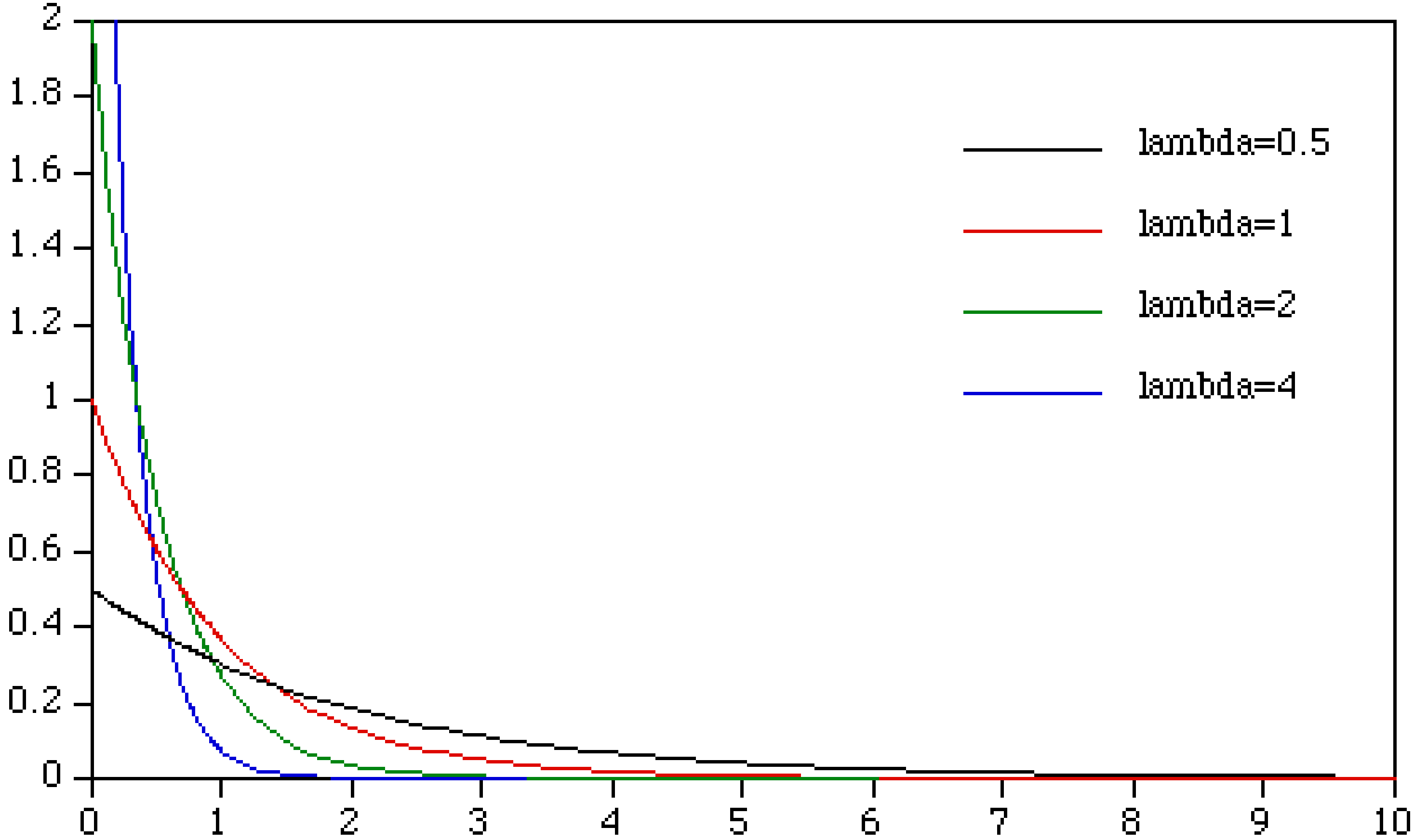
$$E(x) = 1/\lambda \qquad \mathrm{Var}(x) = 1/\lambda^2$$

Examples: Exponential

- Lifetime of electronic components
- Waiting times in a queue
- Time between successive random events (e.g. earthquakes, eddies – testing for randomness)
- Also used to model distances:
  - distance between mutations on a DNA strand
  - Distance between scallops in a scallop bed

MSL
FISH 604

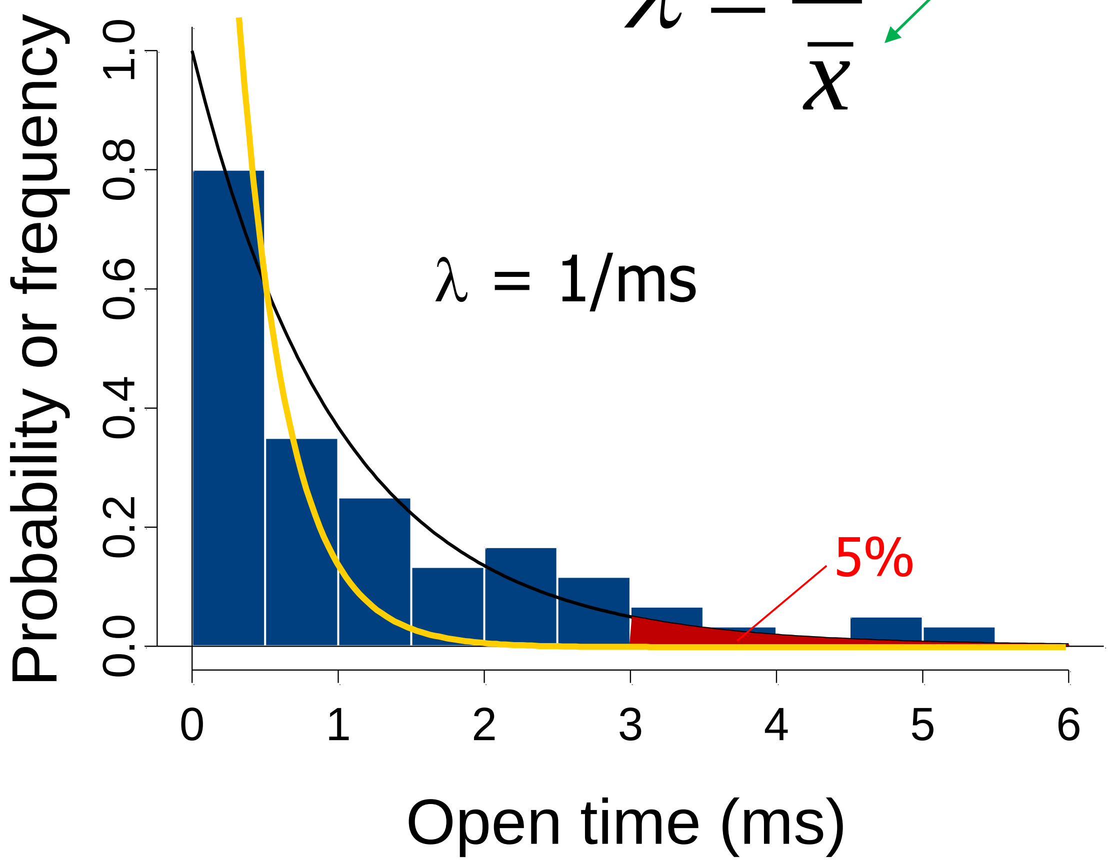
Exponential distributions



Exponential distribution

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

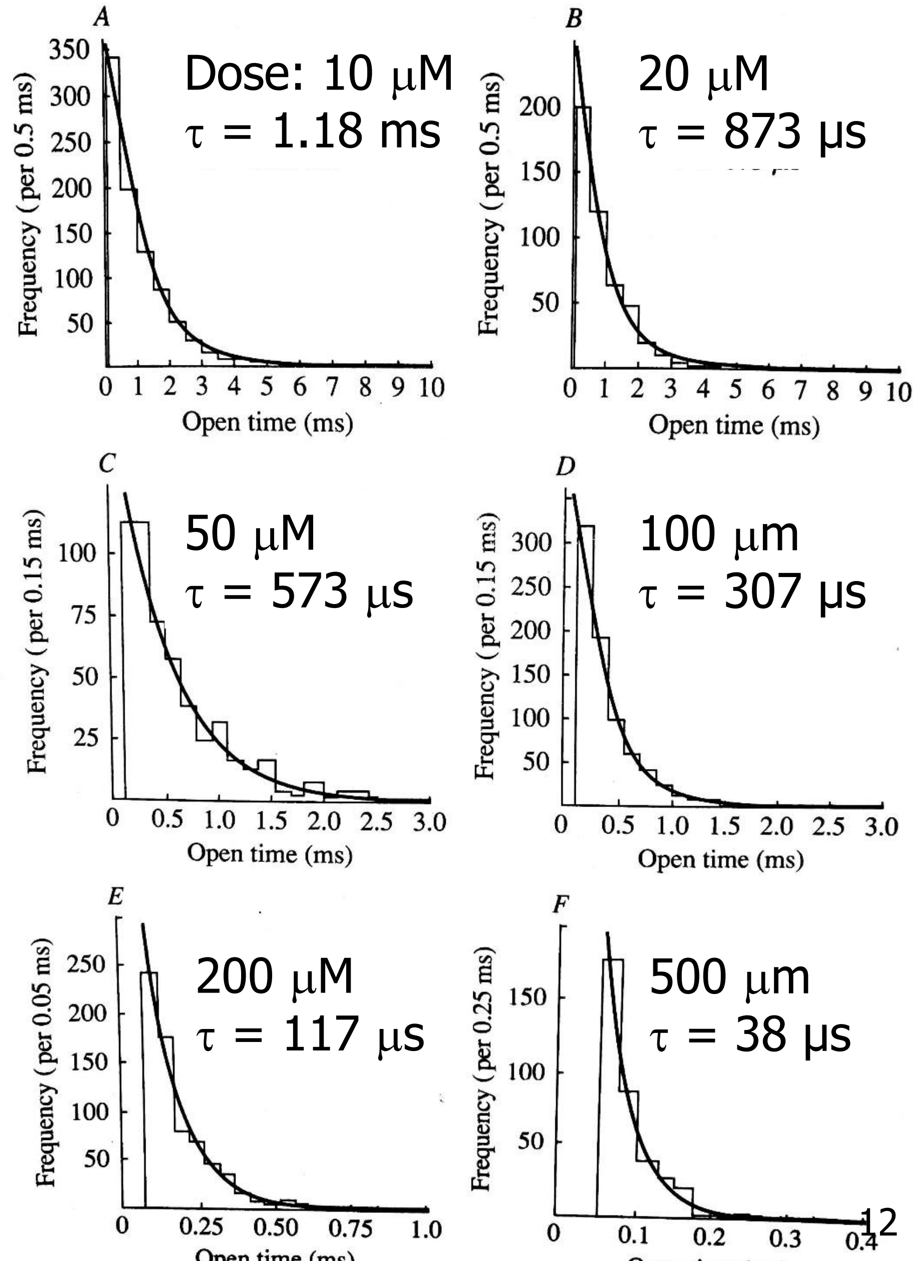
= mean 'open time'



- Sum of open times for all ion channels in a nerve cell
- Random process with an average rate of λ=1 per unit time (ms)
- What are the chances that channels are open more than 3 ms?
- How do drugs affect open times?

Exponential distribution

Histograms of open times at varying concentrations of a channel blocking agent (suxamethonium) with fitted exponential densities (where the rate parameter $\lambda = 1/\tau$)

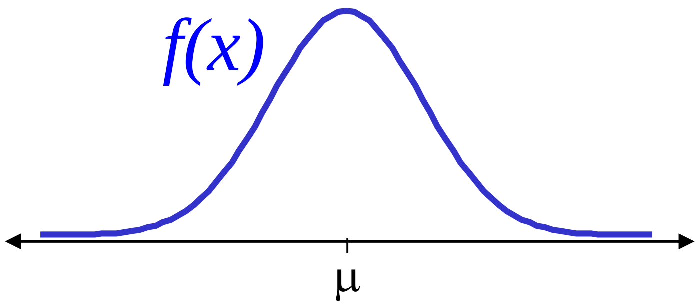


From Rice (1995)

Normal (Gaussian)

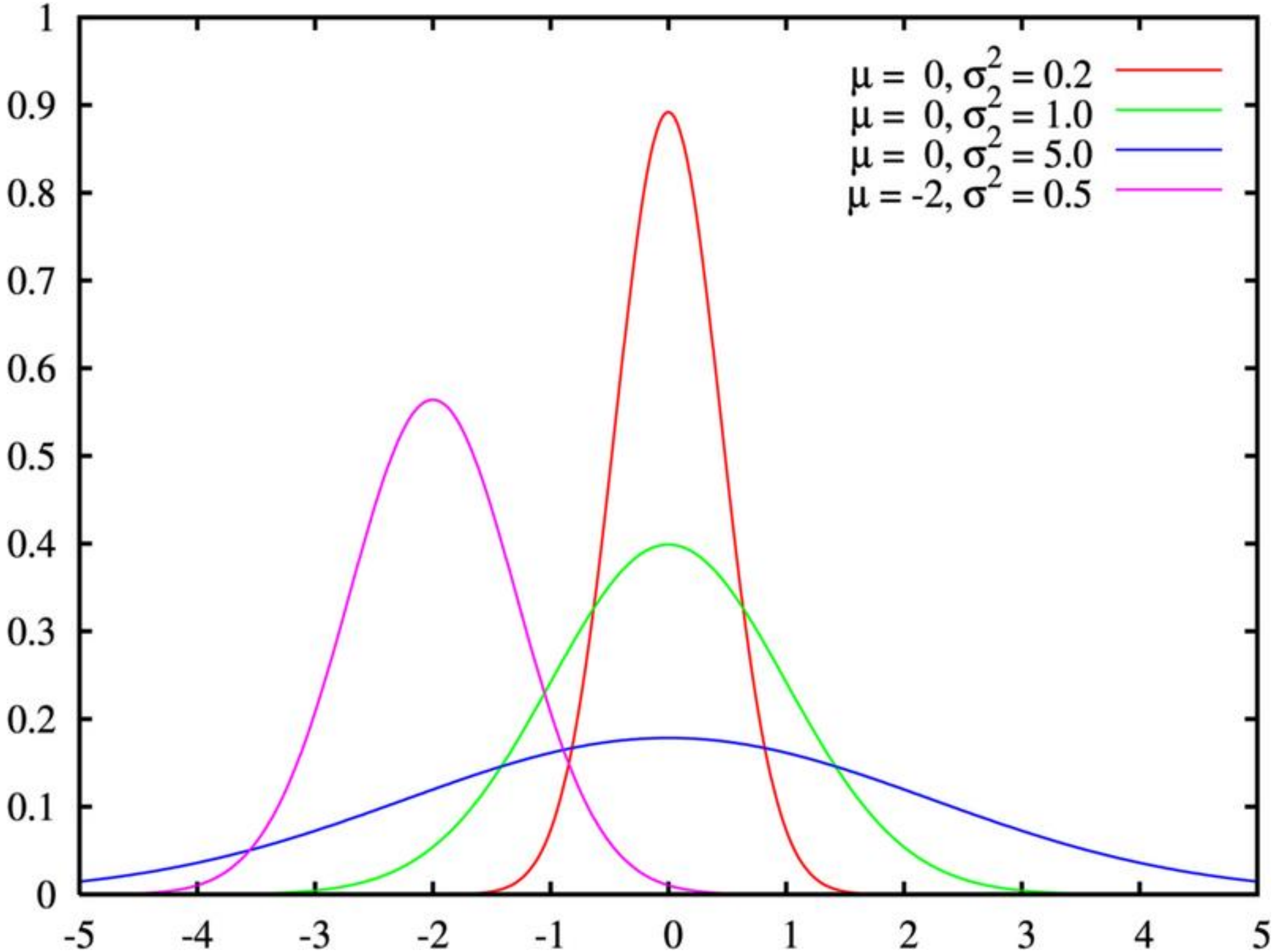
- Familiar "bell-shaped" curve
- Specified by two parameters, the mean (μ) and the variance (σ^2)
- Probability density function:

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)}$$



- Mean: $\boxed{E(x) = \mu}$ Variance: $\boxed{Var(x) = \sigma^2}$

Normal distributions

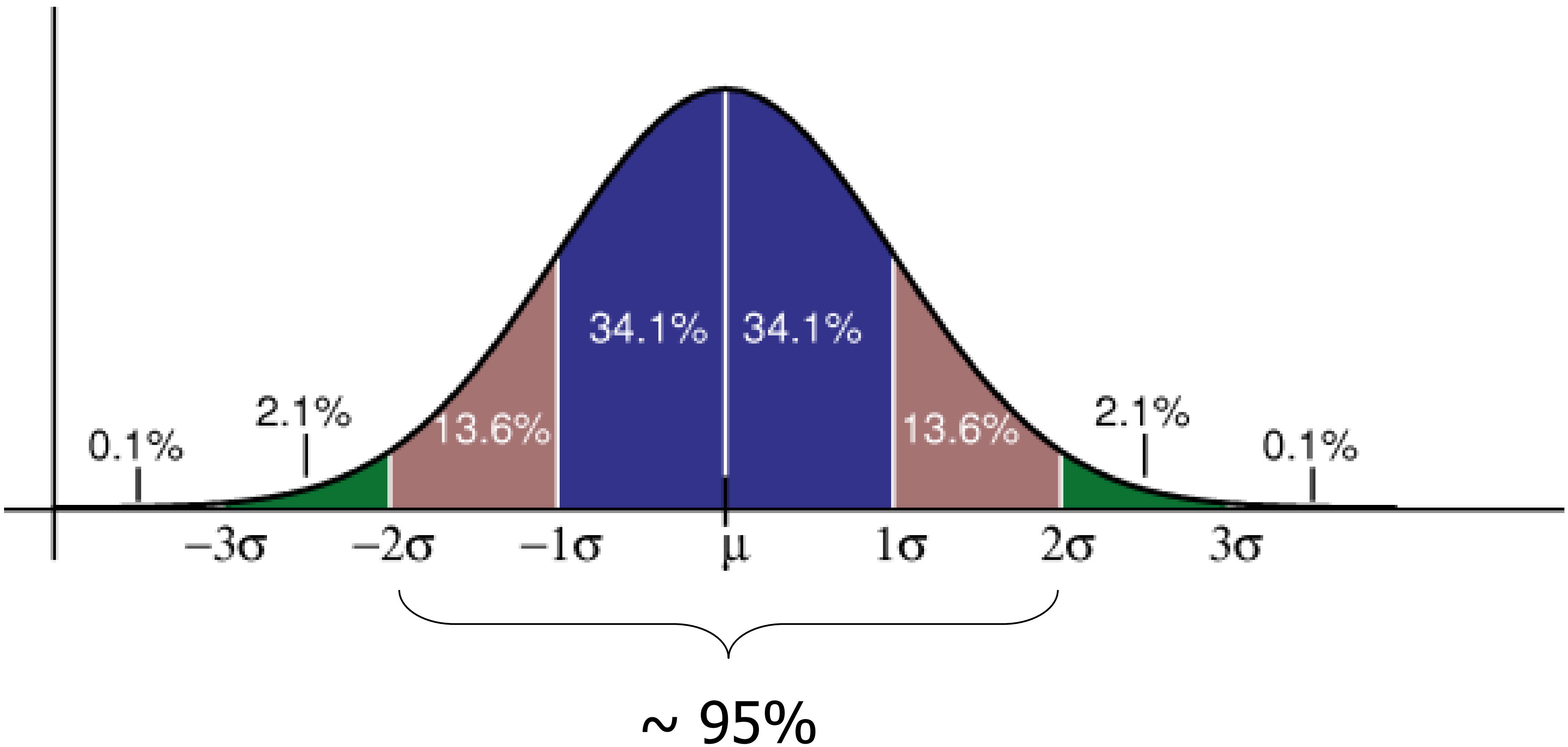


Normal distribution

Why is it so ubiquitous?

- Measurement errors tend to follow a normal distribution (Gauss, Laplace)
- Data on heights and weights of human and animal populations tend to be normally distributed (Quetelet & Galton)
- **Central Limit Theorem**: Sum (or average) of a large number of independent random variables is approximately normally distributed
  - E.g. as *n* increases in the binomial distribution, the sum of outcomes approaches a normal distribution

Normal distribution



Exact 95% confidence interval: $\boxed{\mu \pm 1.96 * \sigma}$

MSL FISH 604

Normal distribution in models

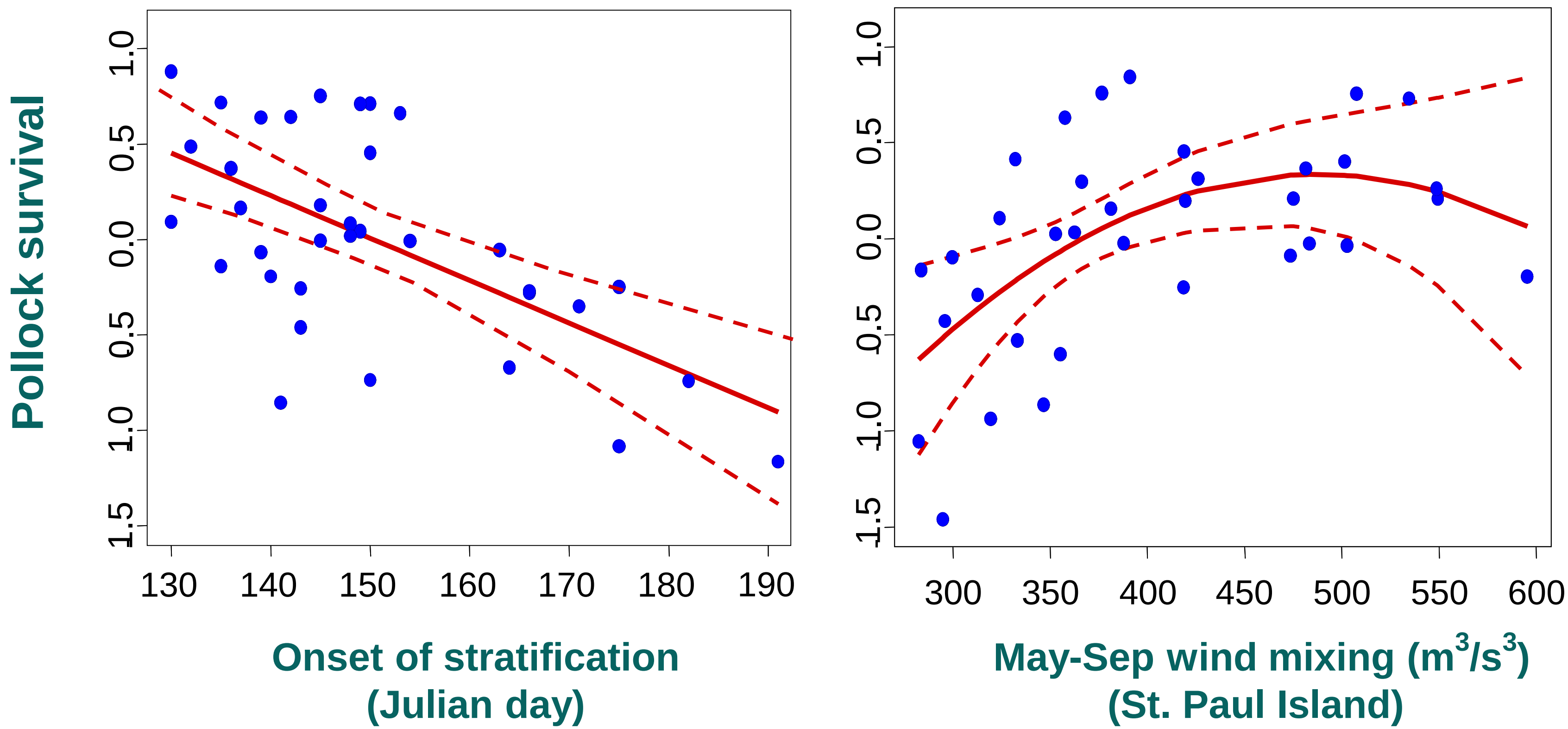
- Most commonly used distribution for residuals in linear and non-linear models:

$$y = f(X, \theta) + \varepsilon$$

where: $\varepsilon \sim N(0,\sigma^2)$

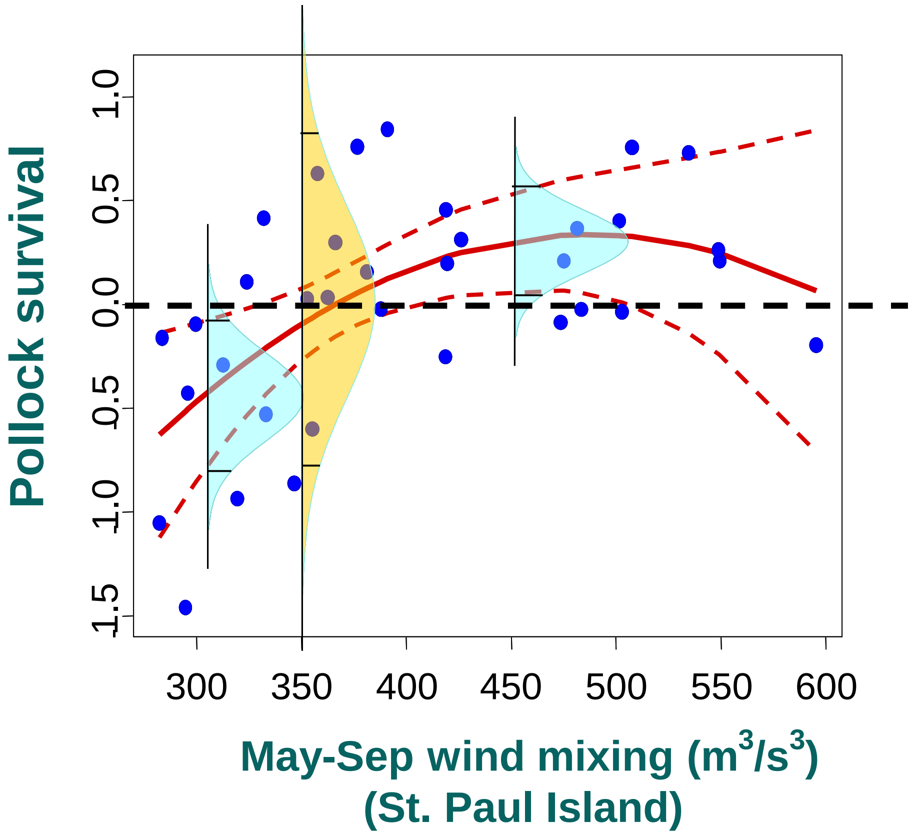
Remember: It's not the data (y), but the error (ε) that is assumed to be normally distributed!!

Pollock example



From: Mueter et al. (2006)

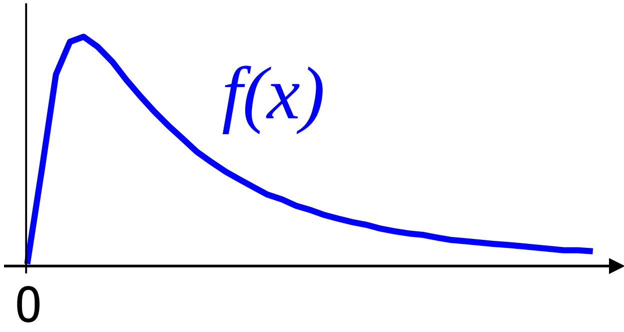
Pollock example



Log-normal distribution

- Probability of a random variable whose logarithm is normally distributed
- Probability density function:

$$\boxed{f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right)}$$

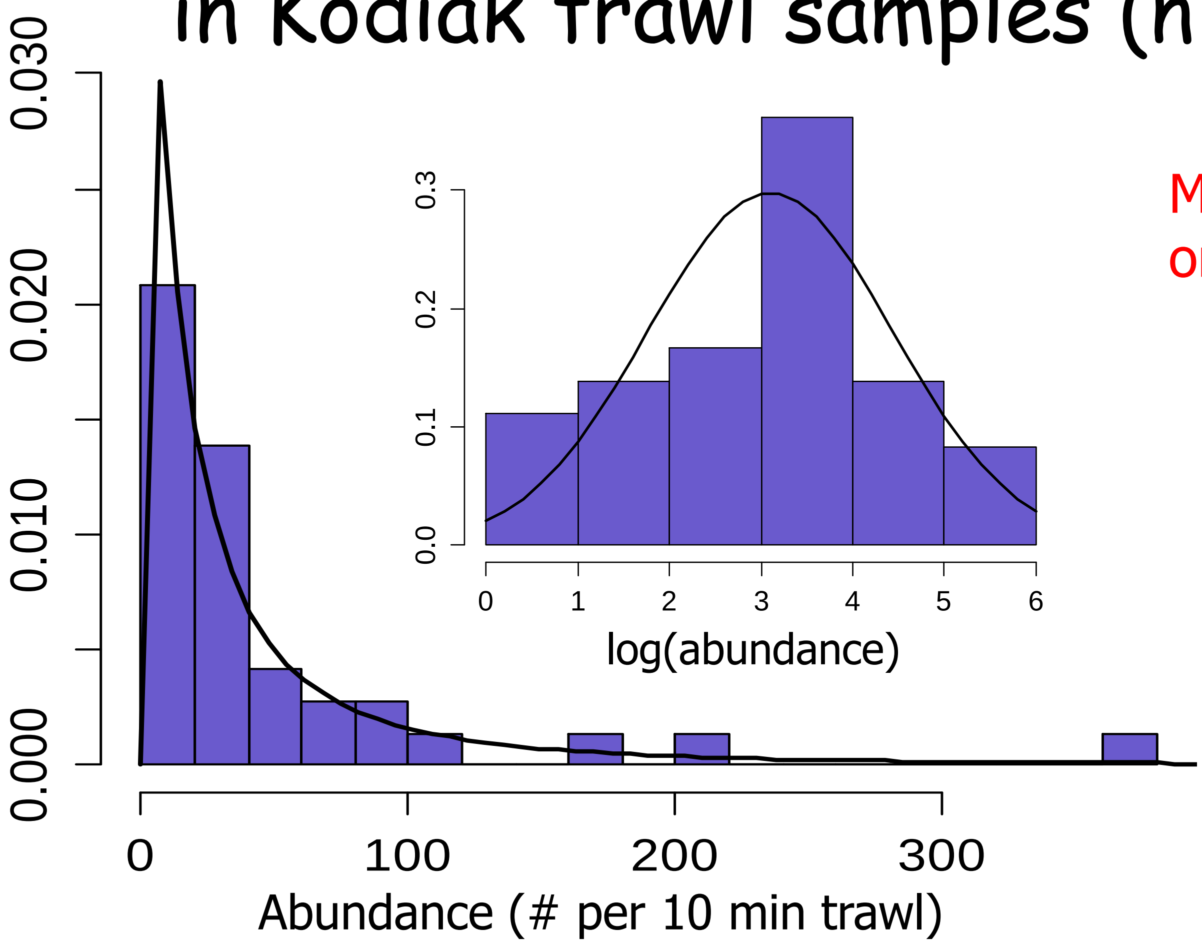


- Mean: $\boxed{E(x) = e^{\mu+\sigma^2/2}}$

- Variance: $\boxed{Var(x) = (e^{\sigma^2} - 1) \cdot e^{2\mu+\sigma^2}}$

Examples: Log-normal

Abundances (where present) of jellyfish in Kodiak trawl samples (n = 36):



Mean and variance on log-scale:

$$\hat{\mu} = 3.10$$

$$\hat{\sigma}^2 = 1.79$$

<u>Exercise</u>: Compute mean and variance on original scale!

MSL FISH 604

Gamma distribution

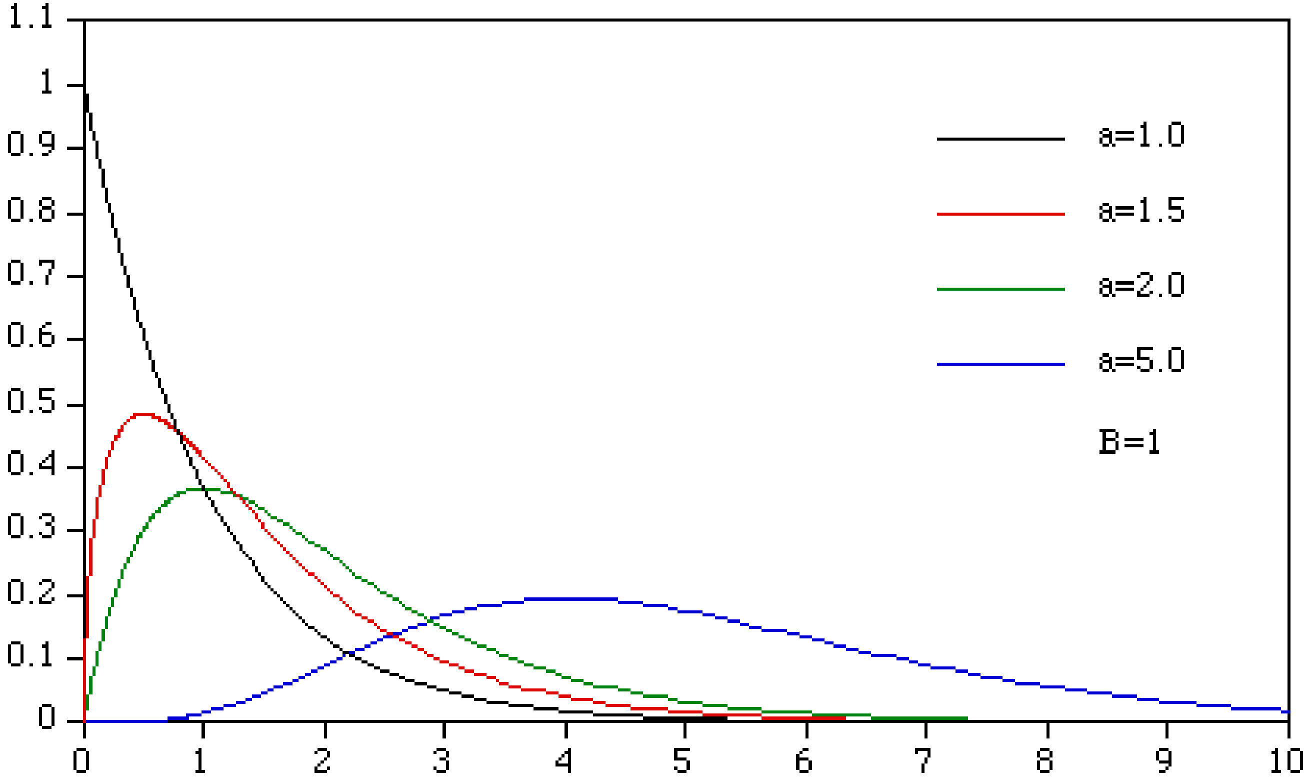
- Flexible class for modeling non-negative random numbers, characterized by a shape (α) and a rate parameter (β):

$$\boxed{f(x) = x^{\alpha-1}\frac{\beta^{\alpha}e^{-\beta x}}{\Gamma(\alpha)} \qquad \text{for } x > 0}$$

- Mean: $\boxed{E(x) = \alpha/\beta}$

- Variance: $\boxed{Var(x) = \alpha/\beta^2}$

Gamma distributions



MSL FISH 604

Examples: Gamma

- Gamma is a generalization of the exponential; it describes the waiting time until the r^{th} event for a process that occurs randomly over time at a rate of β
- Used to model patterns of occurrence of earthquakes (in time, in space, magnitude) (more flexible than exponential)
- Distribution of lifetimes
- **Abundance of animals in random samples (similar to log-normal)**

Further reading

MSL FISH 604

Assigned reading:

- Gotelli, N.J., and Ellison, A.M. 2004. A Primer of Ecological Statistics. (Chapter 2)

Additional reading:

- Balakrishnan, N., 2003. A primer on statistical distributions. Wiley-Interscience, NY.
- Evans, M., Hastings, N., and Peacock, B. 2000. Statistical Distributions. Wiley Series in Probability and Statistics. John Wiley & Sons.