# MSL / FISH 604
## *Module 2 (part 2) :*
## Basic statistical concepts: Distributions

Instructor: Franz Mueter
Lena Point, Rm 315
796-5448
fmueter@alaska.edu

MSL FISH 604

# Objective and outcomes

## Objectives

- Review / introduce discrete probability distributions

## Learning outcomes

- Know the major discrete 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

# Review & preview, Module 2: Basic statistical concepts

MSL FISH 604

- Data summaries
  - Location: Mean, median, quantiles
  - Spread: Variance, Standard deviation, MAD, IQR
  - Graphical summaries
  - Expectation and variance; variance estimation
- Probability and probability distributions
- Distributions
  - Discrete (Binomial, Multinomial, Poisson)
  - Continuous (Uniform, Exponential, Normal or Gaussian, Log-normal, Gamma)

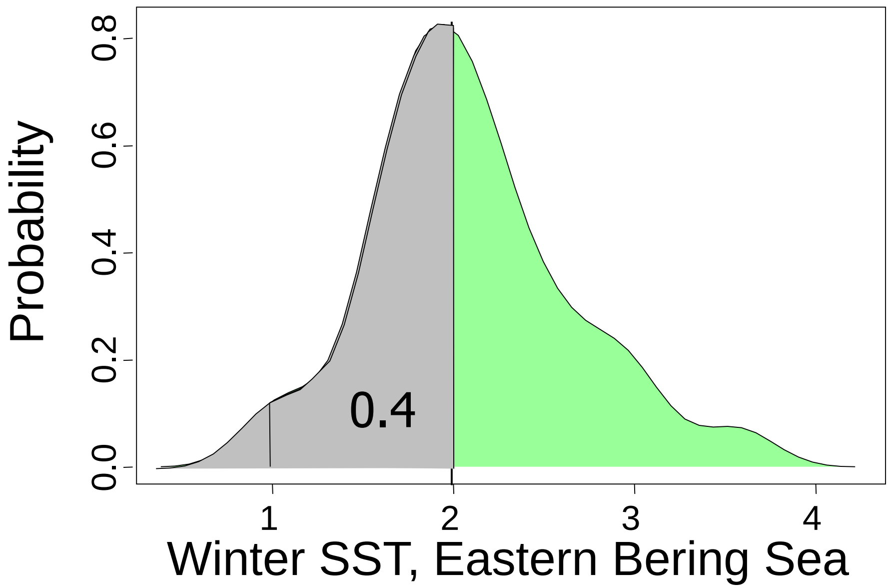
Today

Next time

# Quantifying probability

- To quantify probabilities in statistics we generally use **probability distributions (=probability frequency OR probability density function: pdf)**, which assign a probability to each value of a quantity of interest (or to a range of values, in the case of continuous variables)
- **Example:** Using probabilities to describe climate state of the eastern Bering Sea

| State | *p* |
|---|---|
| Cold | 0.2 |
| Average | 0.5 |
| Warm | 0.3 |



MSL FISH 604

# Why probability distributions?

- Testing hypotheses
  - Distribution of test statistic ($T_{obs}$)
  - Quantify probability that $T_{obs} > T_{crit}$
- Fitting models to data
  - Quantify probability (likelihood) of data, given a known distribution and a set of parameters
- Quantify uncertainty: variance, confidence intervals, or <u>full probability distribution</u> of:
  - Observations
  - Predictions
  - Parameters

MSL
FISH 604

# Some distributions

- Discrete:
  - Bernoulli
  - Binomial
  - Multinomial
  - Poisson
  - Negative binomial
- Continuous:
  - Uniform
  - Exponential
  - Normal or Gaussian
  - Log-normal
  - Gamma

# Distributions: Bernoulli

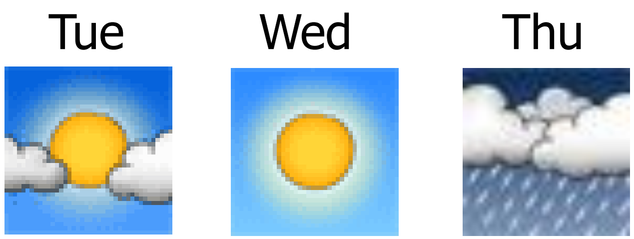
- Probability of a single event
  - Event happens: Probability $p$
  - Event does not happen: Probability 1-$p$
- Typically denoted by 0/1:

$$\Pr(X = 1) = p$$
$$\Pr(X = 0) = 1-p$$

- Parameter $p$ specifies all there is to know about the event!

$$\boxed{E(X) = p \qquad \mathrm{Var}(X) = p(1-p)}$$

# Example: Probability of rain




- Probability that it rains on a given day: $p$
- How to estimate $p$:
  - based on past frequency:
    $\hat{p}$ = k/N = rainy days/total days
  - Based on simple or complex model:

    $\hat{p} = f$ (winds, temp, etc)

# Binomial

- How often does a particular "event" occur in each of *n* "experiments" or samples (Or: Number of "successes" in *n* Bernoulli "trials")
- Probability distribution: ($k$ = 0,1,2,…)

  where *k* is the number of "successes", *p* is the probability of "success" and *n* is the number of "trials"
- Mean and variance:

# Binomial

- How often does a particular "event" occur in each of *n* "experiments" or samples (Or: Number of "successes" in *n* Bernoulli "trials")
- Probability distribution: ($k$ = 0,1,2,…)

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{where} \binom{\mathrm{n}}{\mathrm{k}} = \frac{n!}{k!(n-k)!}$$

where $k$ is the number of "successes", $p$ is the probability of "success" and $n$ is the number of "trials"

- Mean and variance:

$$E(X) = np \qquad \mathrm{Var}(X) = np(1-p)$$

# Binomial

MSL FISH 604

One possible sequence of 4 'heads' & 5 'tails' in n = 9 trials:

| 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | |
|---|---|---|---|---|---|---|---|---|---|
| q * | p * | p * | q * | p * | q * | q * | q * | p | $= p^k q^{n-k}$ |

Associated probabilities (assuming <u>independent</u> events)

There are $\binom{n}{k}$ ways to get 4 heads in 9 trials!

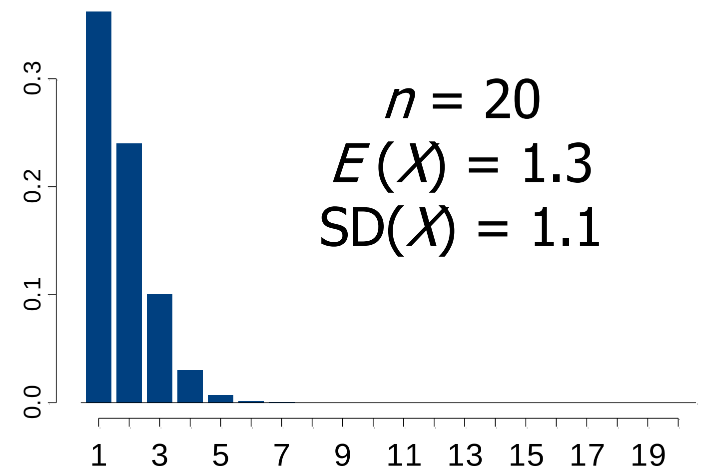
Hence: $$\Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

# Examples: binomial

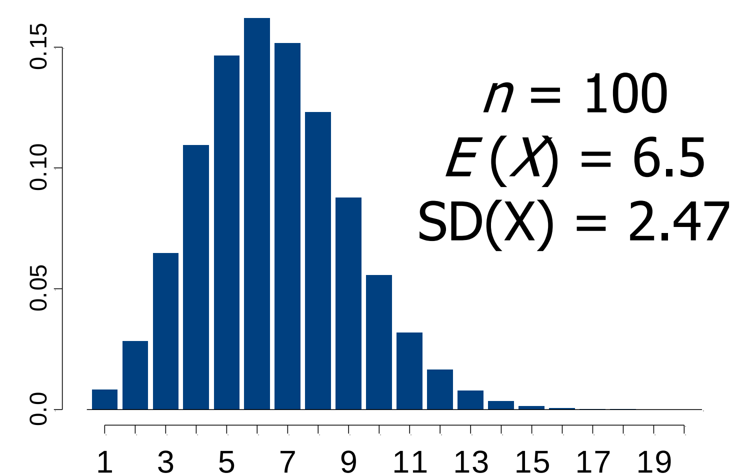
- # of heads in *n* coin tosses!
- # of females in a sample of *n* fish
- # of phytoplankton samples in which diatoms dominate
- Presence/absence of a species (# of samples in which it occurs)
- # of above-average SST years

# Example: binomial

- Sample of $N$ = 46 herring collected after PWS oil spill
- $j$ = 3 herring were deformed
- What is the probability distribution of deformed herring in a sample of 20? A sample of 100?
- First, estimate $p$
  - Based on observed frequency: $\hat{p} = j/N$
  - Based on more complex model: $\hat{p} = f$ (location, sex, age, month)
- Calculate Pr($k$) for k = 0,1,2,... assuming binomial distribution and plot probabilities against k

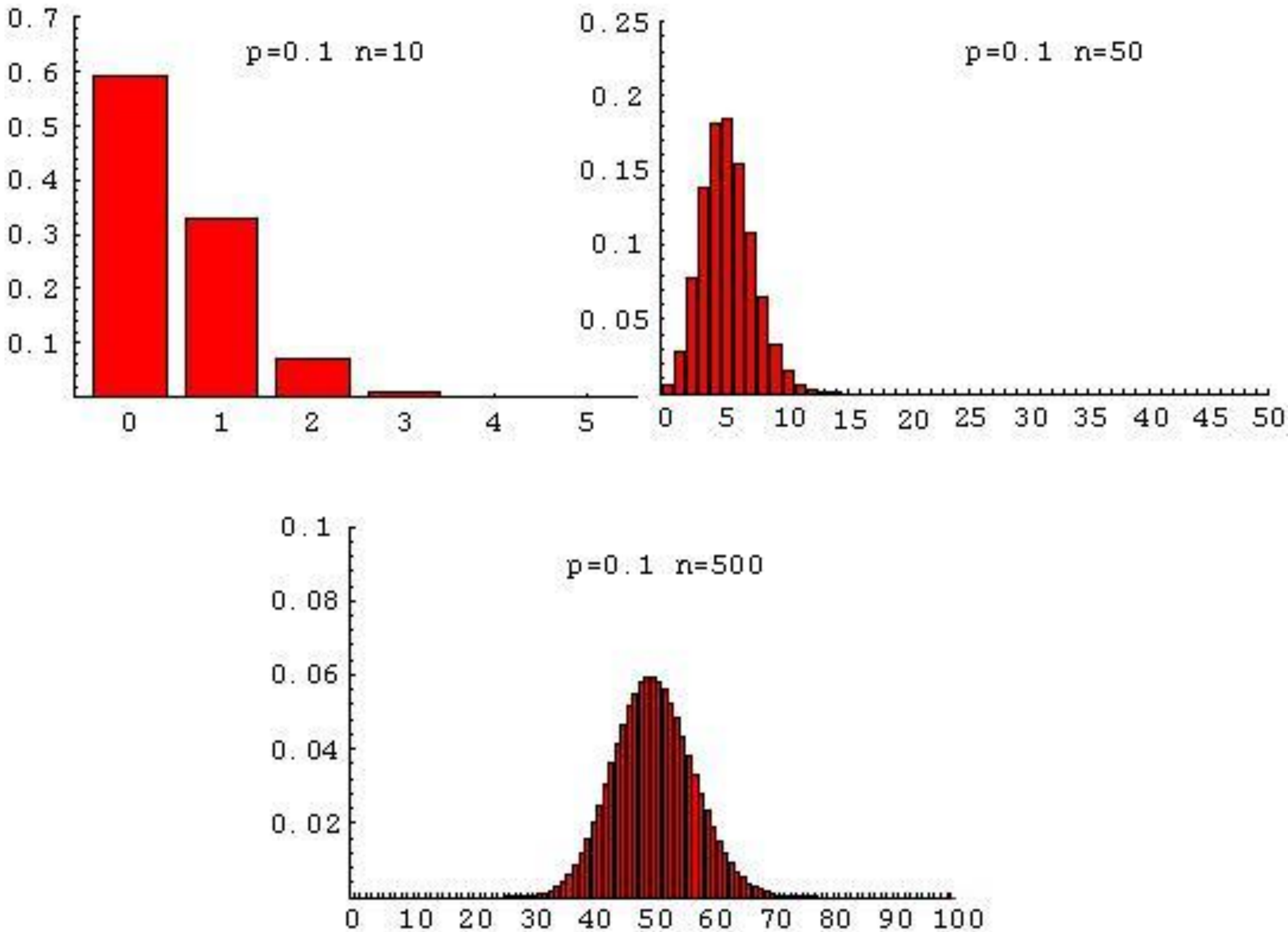



```
> barplot(dbinom(1:20,20,3/46))
```




```
> barplot(dbinom(1:20,100,3/46))
```

# Normal approximation



MSL
FISH 604

# Multinomial

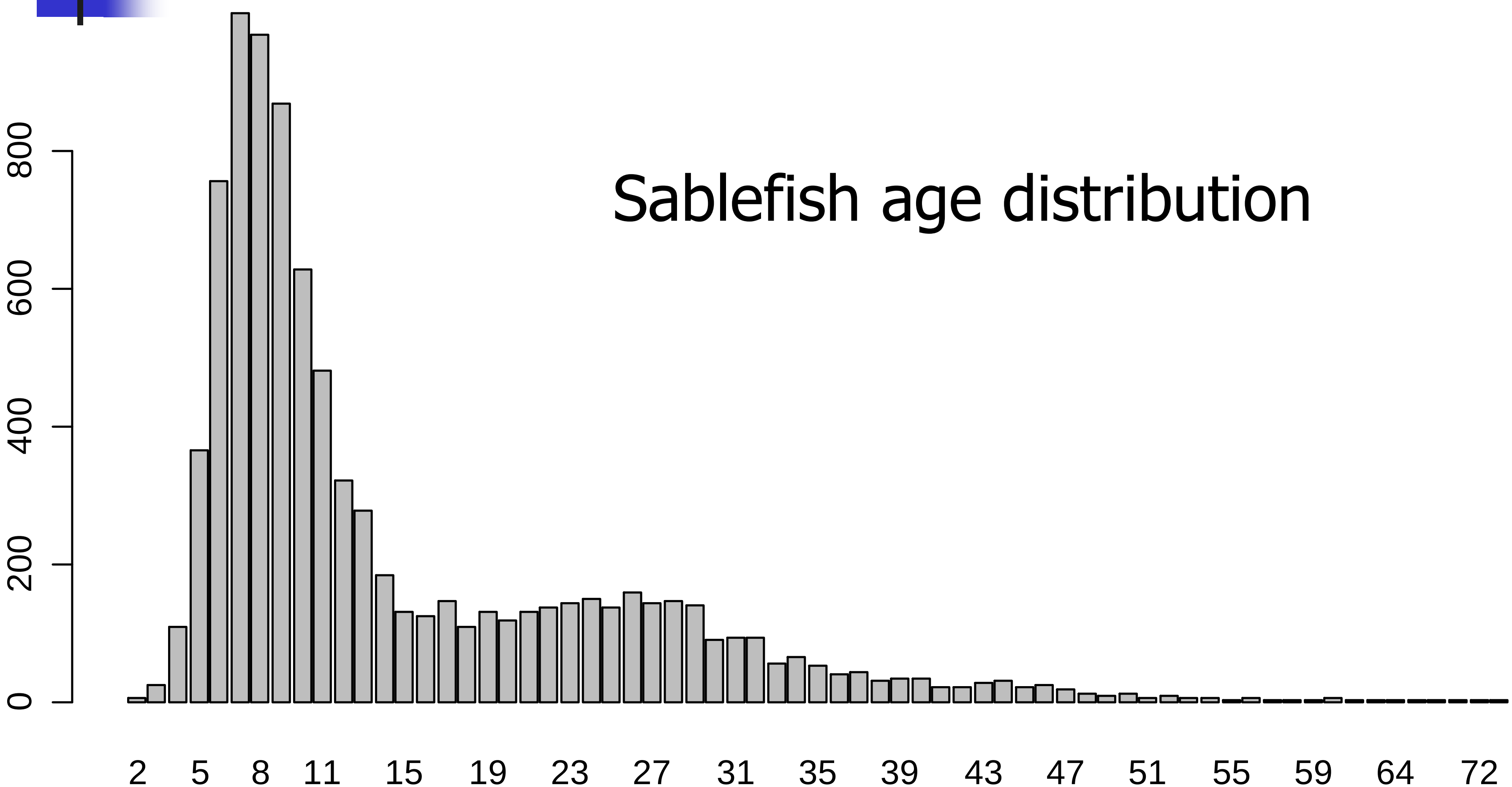
- Generalization of the binomial distribution with more than two outcomes ($r$ "classes")
- Probability of observing $n_i$ "successes" in the $i$ th class ($i = 1,2, \ldots$r):

$$\Pr(\mathbf{X} = \{n_1, \ldots, n_r\}) = \frac{n!}{n_1! n_2! \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

- Mean & variance:

$$E(n_i) = np_i \qquad \mathrm{Var}(n_i) = np_i(1 - p_i)$$

# Example: Age distribution




Stock assessment models estimate the proportions of fish at a given age and compare **predicted numbers-at-age** to **observed numbers-at-age**

# Example: Salmon ages

- Long-term frequencies suggest that of all sockeye salmon returning to Bristol Bay, 22% are 3 years old, 63% are 4 years old, and 15% are 5 years old

  i.e. $p = \{0.22, 0.63, 0.15\}$
- If we catch 3 fish, what are the possible age distributions and their respective probabilities? (Compute probability for at least one outcome).
- What is the most likely outcome?

# Poisson

- Used for counts of rare events, a single 'rate' parameter ($\lambda$) determines the entire distribution:

$$\boxed{\Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad \mathrm{k} = 0,1,2,\ldots}$$

- Mean and variance

$$\boxed{E(X) = Var(X) = \lambda}$$

- Can be derived as the limit of a binomial distribution with $n \to \infty$ and $p \to 0$, while $np = \lambda$ stays constant!
- Negative binomial similar to Poisson but has extra parameter to allow for "overdispersion"

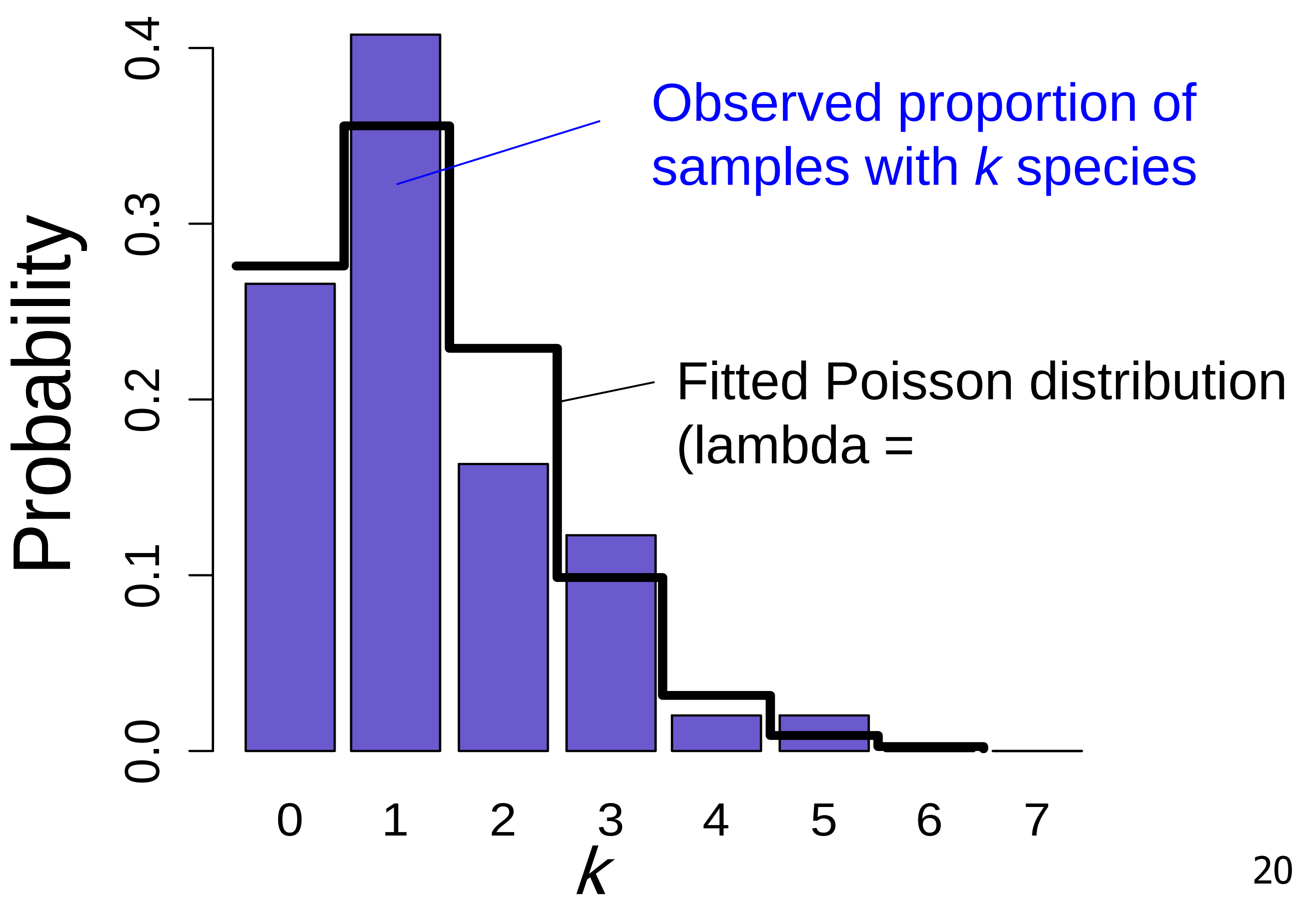
# Examples: Poisson

- Number of radioactive particles emitted from a radioactive source during a period of time
- Insurance companies use it to model number of freak accidents for a large population in a given time period
- Catch of uncommon species in a fishery (# of tuna or sharks per set)
- Count of (rare) species in a sample
- Count of infected organisms (rare disease)

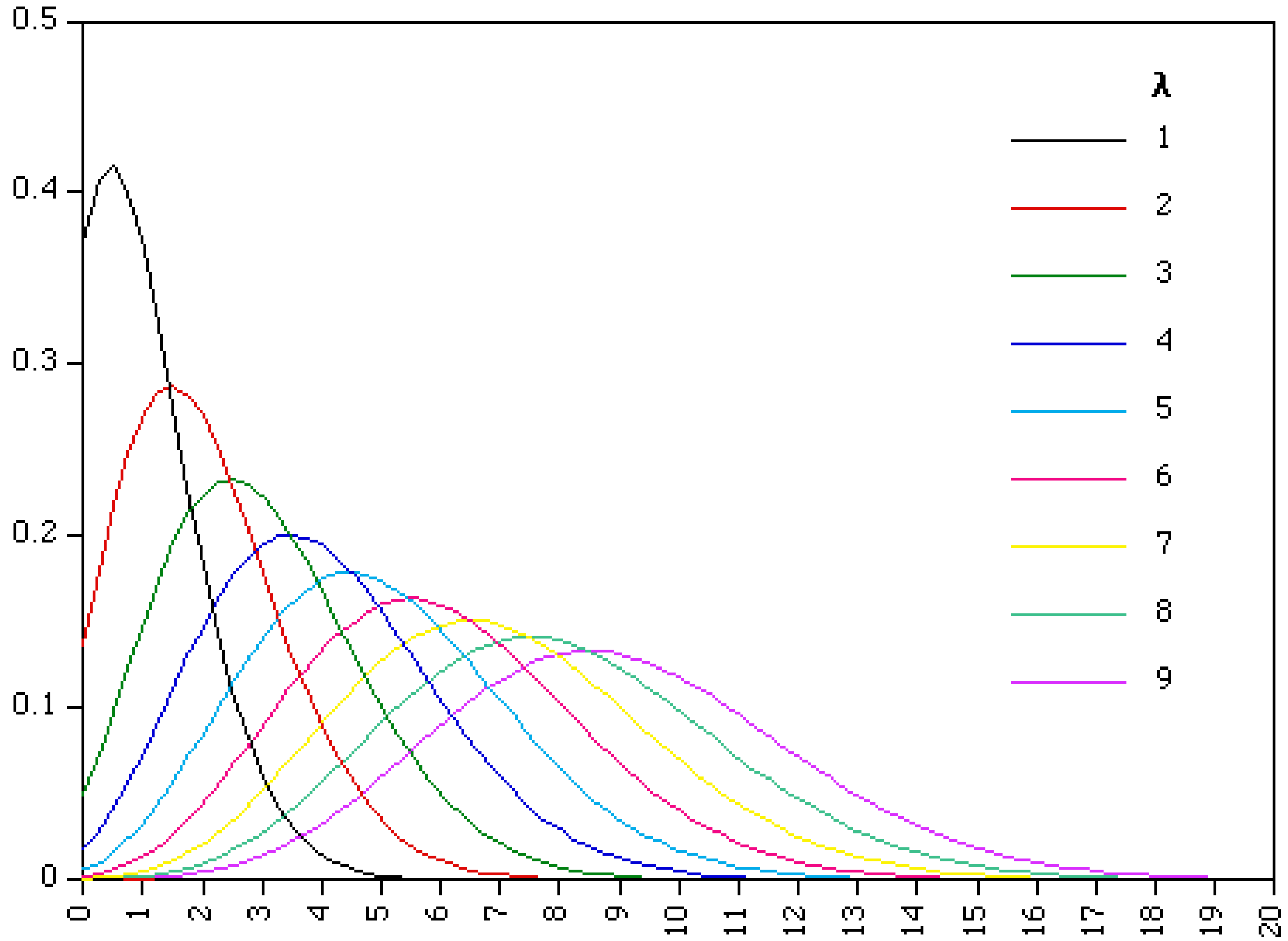
# Example: Poisson

- Number of *Sebastes* species in trawl hauls taken around Kodiak Island

<u>49 hauls total</u>
13 hauls had no Sebastes species
20 hauls had 1 species
8 hauls had 2 species
6 hauls had 3 species
1 haul had 4 species
1 haul had 5 species
(blue bars)



# Poisson distributions



# Negative binomial

- Probability distribution of the number of failures in a sequence of Bernoulli trials needed to get specific number of successes:

$$\Pr(X = k) = \binom{r+k-1}{k} \mathrm{p}^{\mathrm{r}} (1-\mathrm{p})^{\mathrm{k}} \qquad \mathrm{k} = 0,1,2,\ldots$$

- Mean and variance

$$E(X) = r\frac{1-p}{p} \qquad \mathrm{var(X)} = r\frac{1-p}{p^2}$$

# Example: Negative binomial

- Used for count data when variance is larger than Poisson variance (i.e. if variance is larger than mean)
  - For example number of individual fish per trawl sample or other animals with clustered distribution
    (if cluster size follows logarithmic series, total number of individuals follows neg. binomial)

# Further reading



**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.