Generalized Linear Models in Practice

EES 5891-03

Bayesian Statistical Methods

Jonathan Gilligan

Class #14: Tuesday, October 18 2022

Announcements

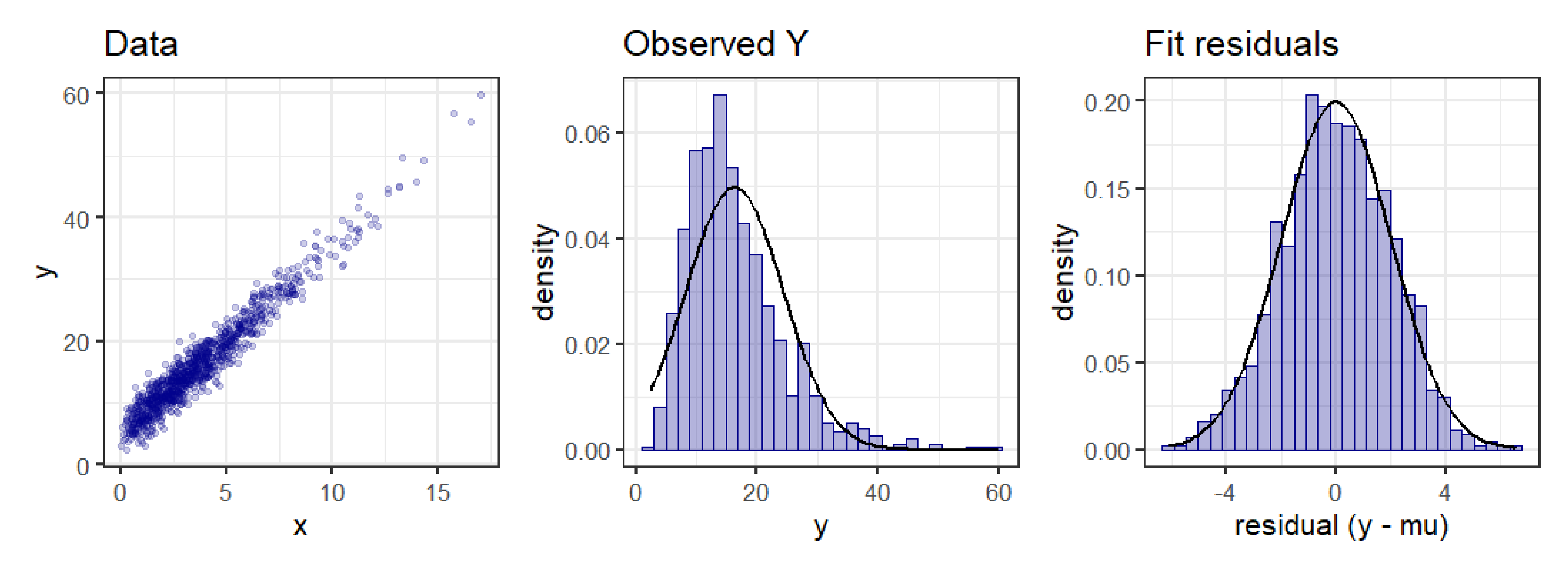
- No office hour this afternoon
- Next homework will be due Tuesday Nov. 1
- Project proposal due Thursday Nov. 3
    - I’ll talk more about the project later today

Choosing Likelihood Functions

Choosing A Distribution

- "Histomancy" (**bad**)
  - Choosing a distribution from a histogram of the outcome variable
  - A Normal distribution means the *residuals* should look Gaussian *after* subtracting the prediction.
- Instead, think about the *process* that produced the observations.

```
a <- 5
b <- 3
sigma <- 2
N <- 1000
d <- tibble(x = rgamma(N, shape = 2, scale = 2),
            mu = a + b * x,
            y = rnorm(N, mu, sigma),
            res = y - mu)

ds <- summarize(d, ymin = min(y), ymax = max(y),
                ymean = mean(y), ysd = sd(y),
                rmin = min(res), rmax = max(res))

ref_norm <- tibble(y = seq(ds$ymin, ds$ymax, length.out = 100),
                   y_dens = dnorm(y, mean = ds$ymean, sd = ds$ysd),
                   res = seq(ds$rmin, ds$rmax, length.out = 100),
                   res_dens = dnorm(res, mean = 0, sd = sigma))
```



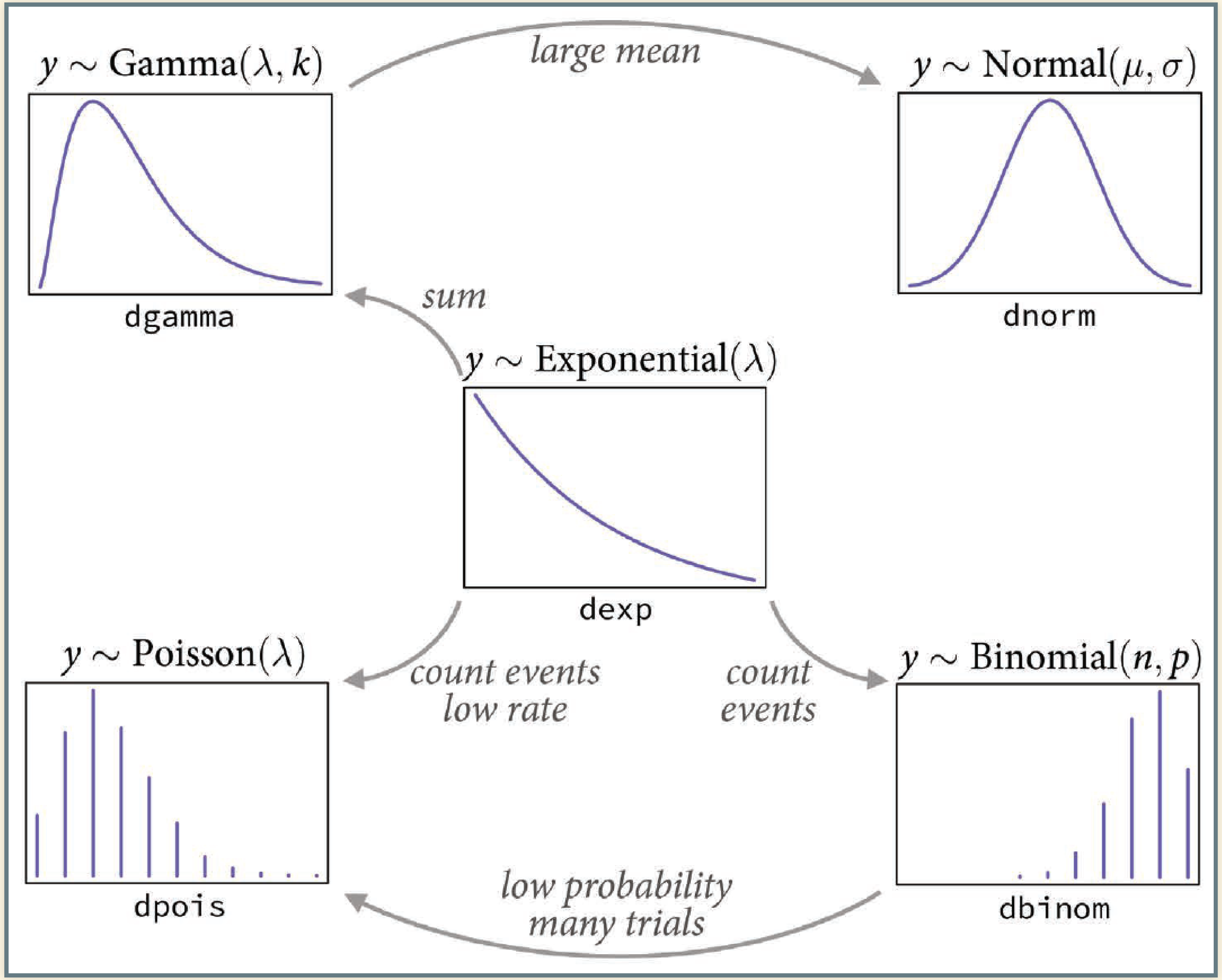
Code for plots

```
library(patchwork)

p1 <- ggplot(d, aes(x = x, y = y)) +
  geom_point(color = "darkblue", alpha = 0.2) +
  labs(title = "Data")

p2 <- ggplot(d, aes(x = y)) +
  geom_histogram(aes(y = after_stat(density)), bins = 30,
                 color = "darkblue", fill = alpha("darkblue", 0.3)) +
  geom_line(data = ref_norm, mapping = aes(x = y, y = y_dens), color = "black",
            size = 1) +
  labs(title = "Observed Y")

p3 <- ggplot(d, aes(x = res)) +
  geom_histogram(aes(y = after_stat(density)), bins = 30,
                 color = "darkblue", fill = alpha("darkblue", 0.3)) +
  geom_line(data = ref_norm, mapping = aes(x = res, y = res_dens), color = "black",
            size = 1) +
  labs(x = "residual (y - mu)", title = "Fit residuals")

p1 | p2 | p3
```

Exponential Distribution

- Events happen randomly at a *constant average rate* in time or space
- Spacing between events follows an *exponential* distribution
- Example:
  - Radionuclides decay at a constant rate
    - The probability that any nucleus decays in the next second is p (assume $p \ll 1$)
  - Consider a large number N of nuclei
    - On average, $r = Np$ decay every second.
  - Consider the time t between successive decays
    - t is distributed exponentially, with probability density `dexp(t, rate = r)` $= r \exp(-rt) = Np \exp(-Npt)$
  - The time between k decays follows the *gamma distribution*: `dgamma(t, rate = r, shape = k)`
    - `dgamma(t, rate = r, shape = 1)` is the same as `dexp(t, rate = t)`.

Exponential Family of Distribuions

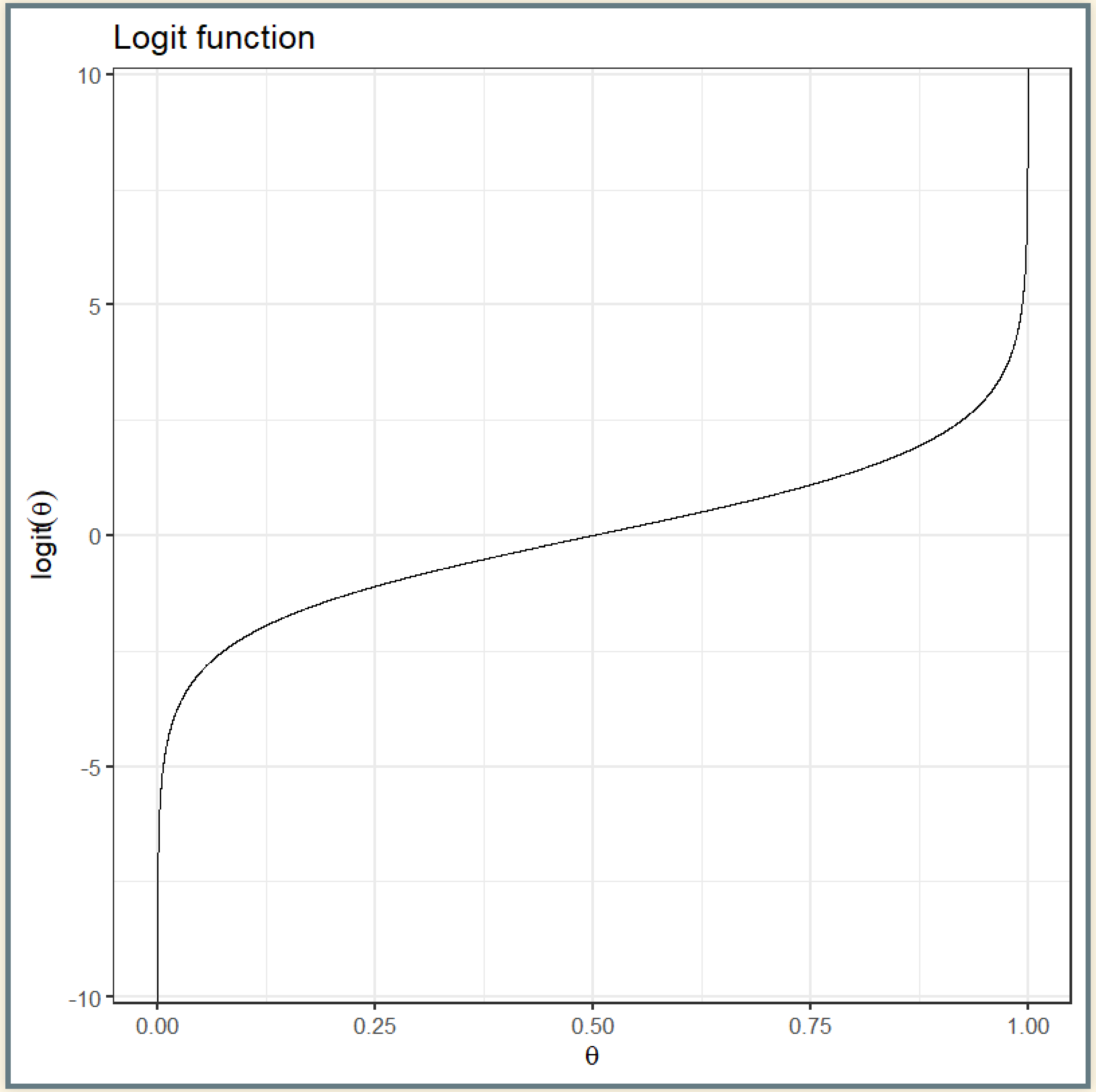


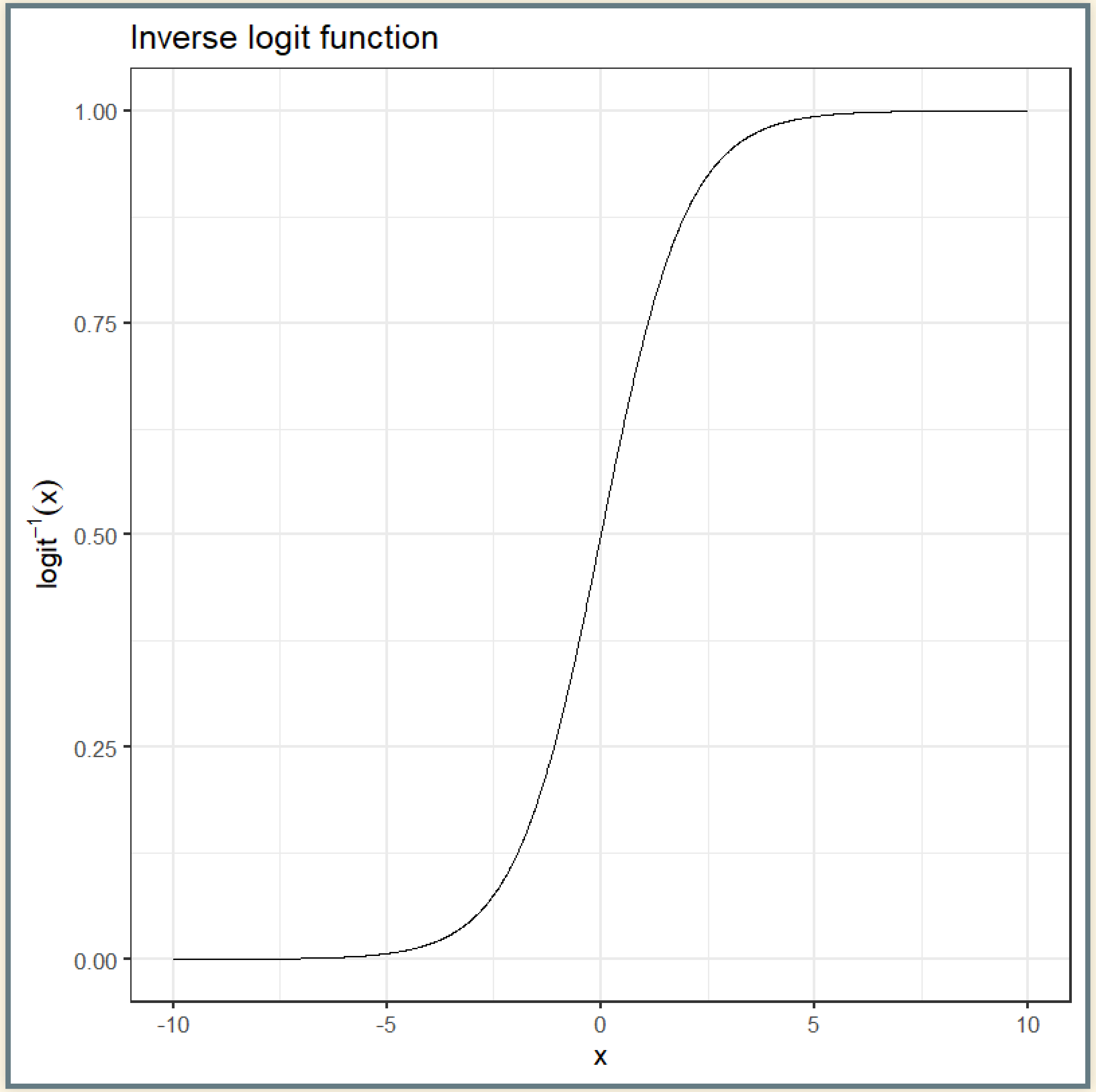
Link Functions

Generalized Linear Models

- Predict y from x: $$\begin{align} y &\sim \text{foo}(\theta, \phi) \\ f(\theta) &= \alpha + \beta x \\ \cdots & \end{align}$$
  - **foo** is some probability distribution with parameters θ and ϕ.
  - $f()$ is a *link function*.
- This is a *generalized linear model* because $f(\theta)$ is a linear function of x.
- Choose a link function based on the constraints on θ.
  - If θ is a probability, then $0 \le \theta \le 1$, so we would need a *link function* to map between the valid range of θ and the unlimited range of $\alpha + \beta x$.
- A common link function for probabilities is the *logit* function $$\begin{align} \text{logit}(\theta) &= \ln \frac{\theta}{1 - \theta} \\ \text{logit}^{-1}(x) &= \frac{1}{1 + \exp(-x)} = \frac{\exp(x)}{\exp(x) + 1} \end{align}$$
  - *logit* maps the range $[0,1]$ to $[-\infty, \infty]$.
  - *logit*$^{-1}$ maps the real axis $[-\infty, \infty]$ to the range $[0,1]$

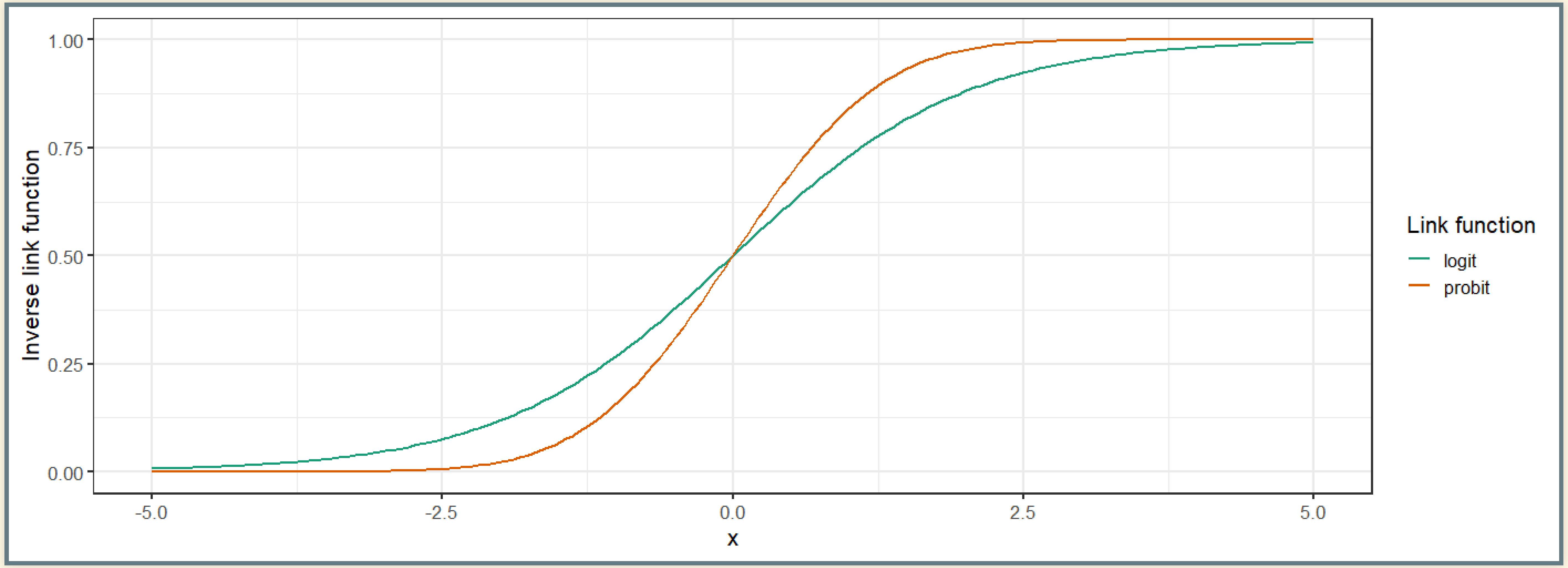
Logit link function





Other Link Functions

- Logit and Probit both map the interval $(0,1)$ to the real number axis $(-\infty, \infty)$. $$\begin{align} \text{logit}^{-1}(x) &= \frac{1}{1 + e^{-x}} \\ \text{probit}^{-1}(x) &= \Phi(x) = \int_{-\infty}^x \text{Normal}(z, 0, 1) dz = \texttt{pnorm(x, 0, 1)} \end{align}$$



Other kinds of link functions

- $\log(\theta) = \alpha + \beta x$ ensures that $\theta > 0$.
  - θ has an exponential dependence on x
    - Constant increments of x cause y to increase by a proportion
      - For $\delta = \log(2) / \beta$, every time $x \rightarrow x + \delta$, $\theta \rightarrow 2\theta$
  - Ask yourself whether this kind of model makes sense, based on what you know about the system you’re studying.

Other considerations

Omitted Variable Bias

- y is 0 or 1
  - True model:
    - Bernoulli model
      - Like coin tossing with probability p of heads

    $$\begin{align} Y &\sim \text{Bernoulli}(p) \\ \text{logit}(p) &= \alpha + \beta_x X + \beta_z Z \end{align}$$
  - Suppose we leave Z out $$\begin{align} Y &\sim \text{Bernoulli}(p) \\ \text{logit}(p) &= \alpha + \beta_x X \end{align}$$
  - With ordinary linear regression, this just increases the scatter around the mean
  - With a logit model, there will be data where X is small, but $Y = 1$
    - These cases will make it seem like X doesn’t have much influence on Y.

Interpreting Parameters

- Compare a linear and a logit model:
  - Model 1: $\mu = \alpha + \beta x$
  - Model 2: $\text{logit}(p) = \alpha + \beta x$
- In Model 1, every time we change x by 1, y changes by β.
- In Model 2:
  - When $\alpha + \beta x$ is close to zero, changing x by δ will change p by $0.25 \beta \delta$,
  - When $\alpha + \beta x$ is close to 3 or -3, changing x by δ will change p by $0.045 \beta \delta$,
  - When $\alpha + \beta x$ is close to 5 or -5, changing x by δ will change p by $0.007 \beta \delta$,

Information Theory

Information Criteria

- Everything from Chapter 7 about information theory applies to GLMs the same as it does to linear models.
- Quetion: Can we apply information criteria like WAIC to decide what likelihood function to use?
  - WAIC and other information criteria only work for comparing models that use the same kind of likelihood function
    - Normal, Binomial, Poisson, etc.
  - Information criteria measure deviance, but the relationship between deviance and how well the model fits the data depends on the likelihood function.
    - If we compare two models with the same likelihood function, smaller deviance means better fit
    - But comparing two models with different likelihood functions, we can't sort out what part of deviance comes from the likelihood and what part comes from goodness of fit.
- Instead, use the **principle of maximum entropy** to choose the likelihood function, and then use WAIC or other information criteria to compare different specifications of the linear part of the model.

Research Project

Research Project

- For most of the rest of the semester, we'll be focusing on research projects
- Choose a statistial problem to work on;
  - Think of:
    - A data set you would like to analyze
    - A question you would like to answer
  - Apply DAG analysis to develop a causal model relating variables
  - Develop a regression model
  - Use your model to analyze your data
  - Develop one or more alternate models and compare them to your original model
  - Intepret what this means for your research question
- On Thursday, we will discuss your ideas for research projects
  - Come to class with thoughts about what you want to work on
  - Formal project proposal is due Thursday Nov. 3