SOCI 620: Quantitative methods 2

Feb 9

Counts and rates

1. Counts as outcomes
2. The Poisson distribution and its assumptions 🐠
3. Log link function
4. Priors through a log link
5. Intercept-only Poisson regression in R
6. Poisson regression with covariates in R

Counts as outcomes

Kinds of counts

Average and deviation

An event that is technically a count, but the scale of process means we can treat it as continuous
E.g. immigration rate, unemployment rate, etc.

Normal distribution

Trials and probability of success

Outcome could have happened at most N times, our data measures how many times it did happen
E.g. “how many days per week…”, etc.

Binomial / Bernoulli distribution

Rate of occurrence

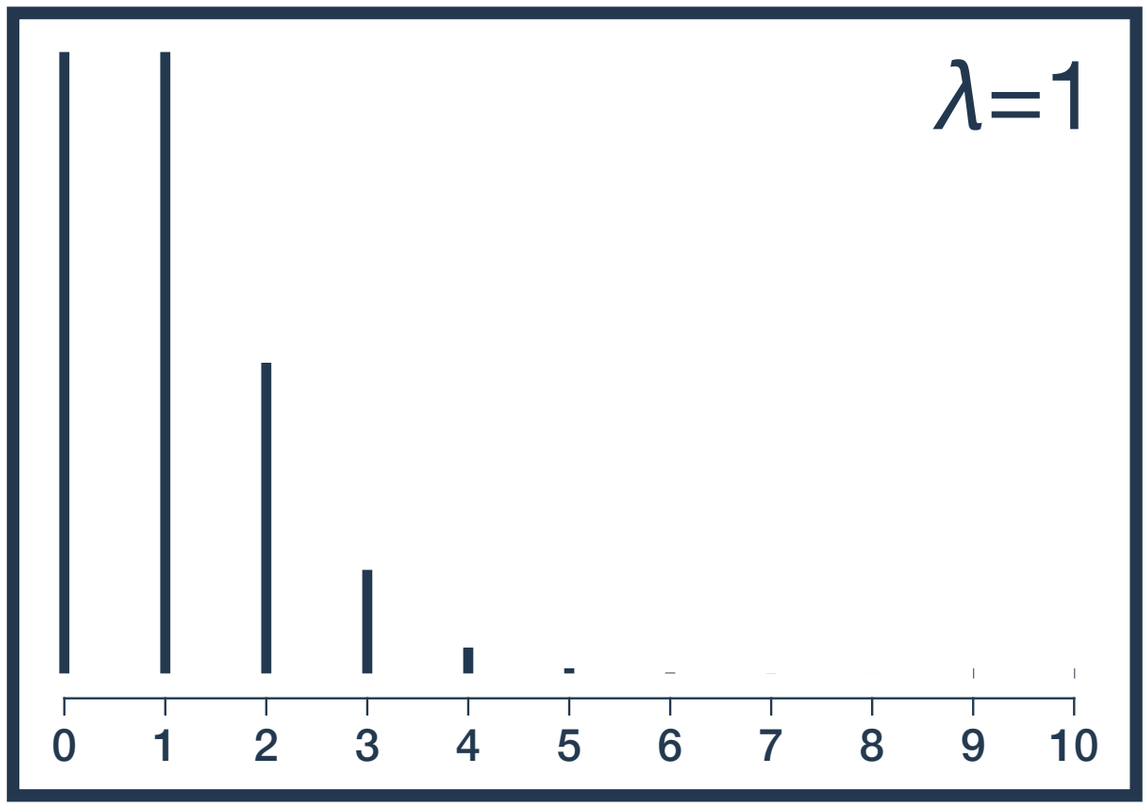
An event that has no (practical) upper limit, but tends to happen at a relatively low rate
E.g. number of friends, grocery stores in a neighborhood, etc.

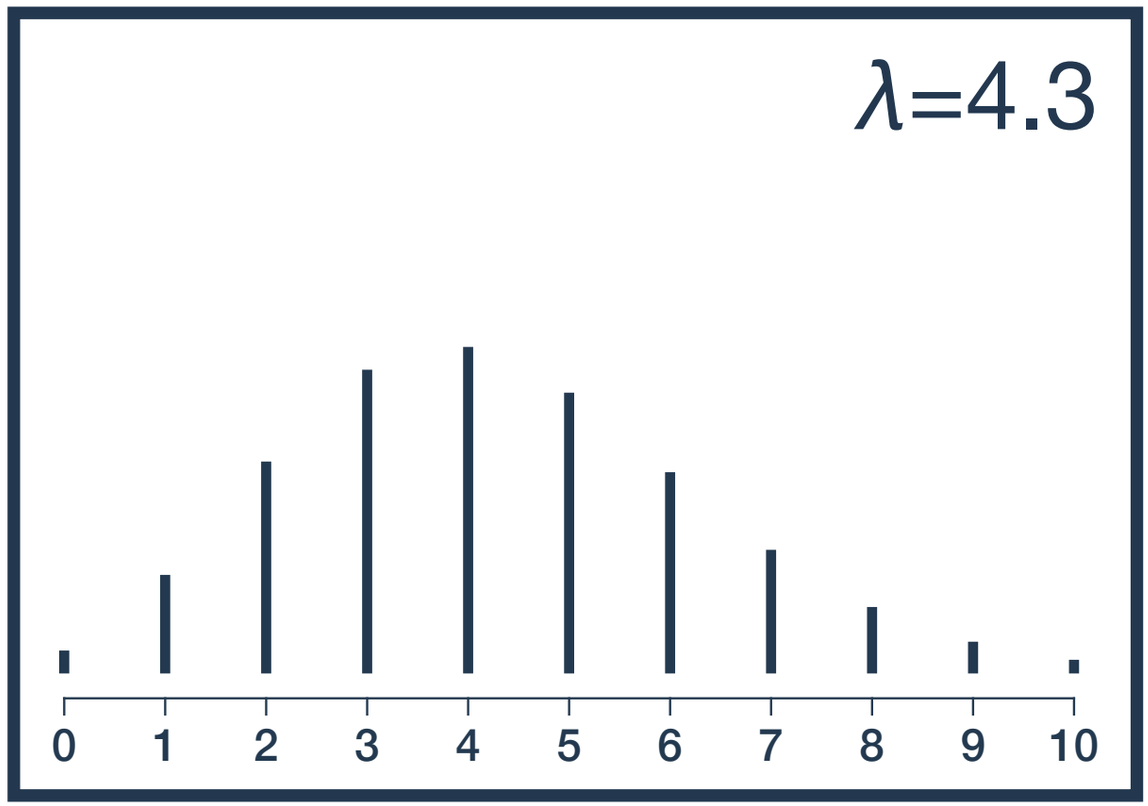
Poisson distribution

Poisson distribution

The *Poisson distribution* gives the probability that an event will happen k times in a particular unit of time or space if it has an average rate of occurrence of λ in that unit of time or space.

$$\text{Prob}(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$



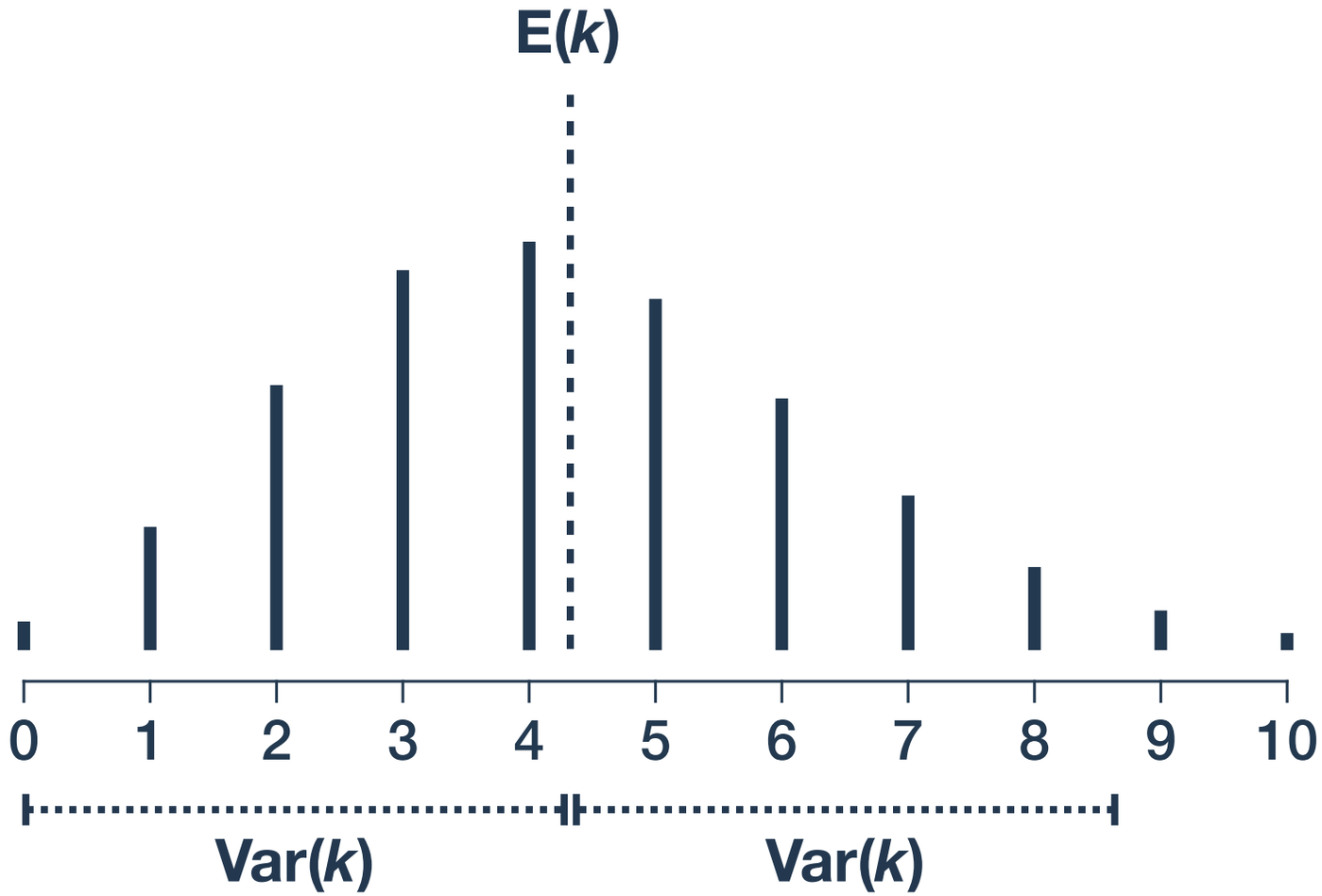


Poisson distribution

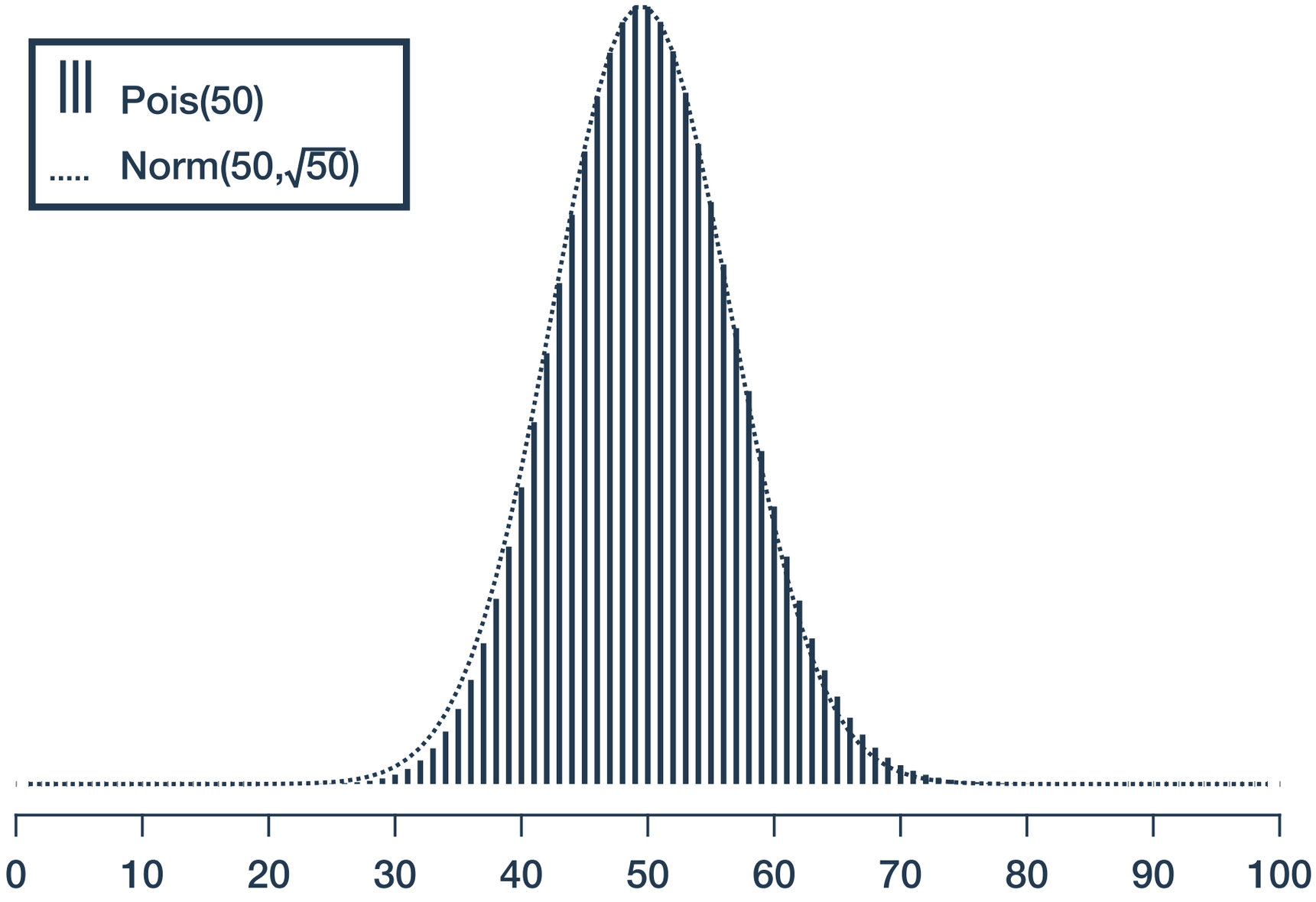
$$k \sim \text{Pois}(\lambda)$$

Single parameter

λ is both the *mean* and the *variance* (s.d. squared) of the Poisson distribution

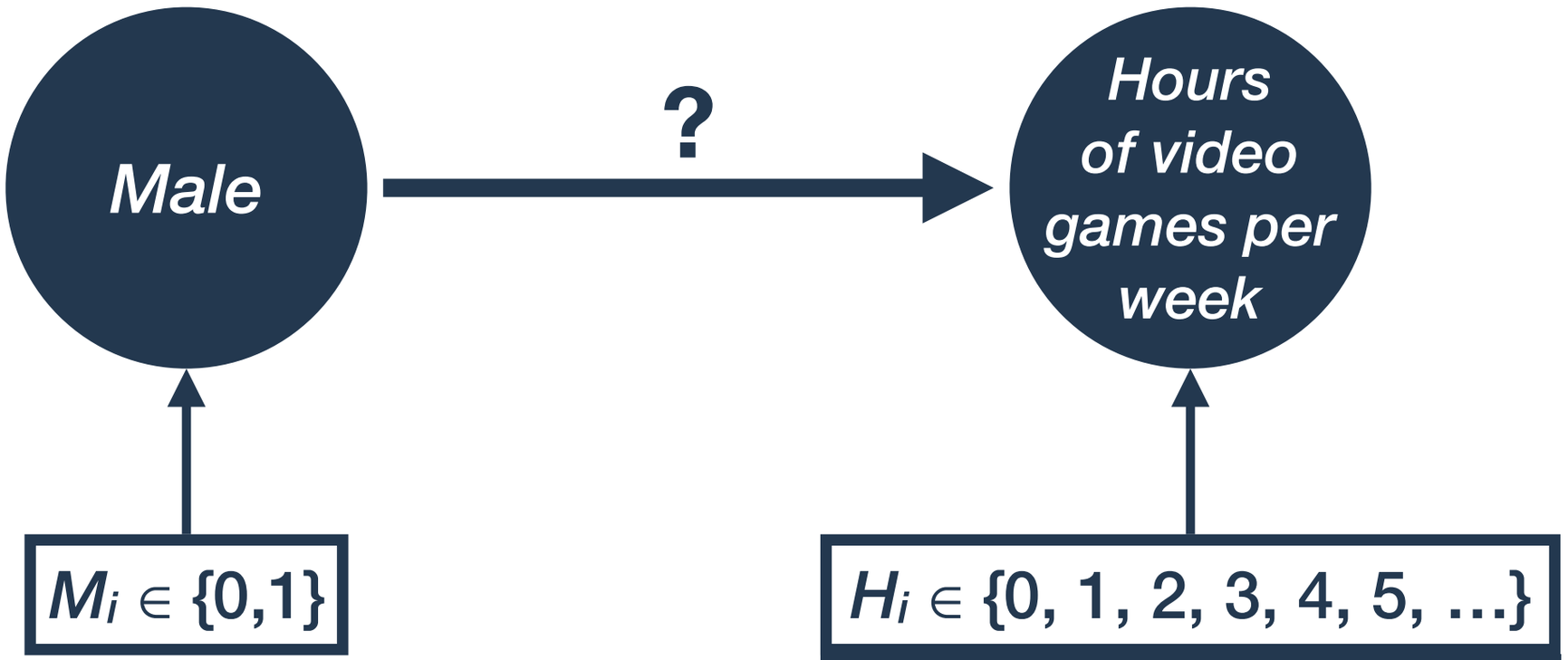


Poisson distribution



Poisson with large λ approximates a normal distribution

Poisson-distributed outcome



$$H_i \sim \text{Pois}(\lambda)$$

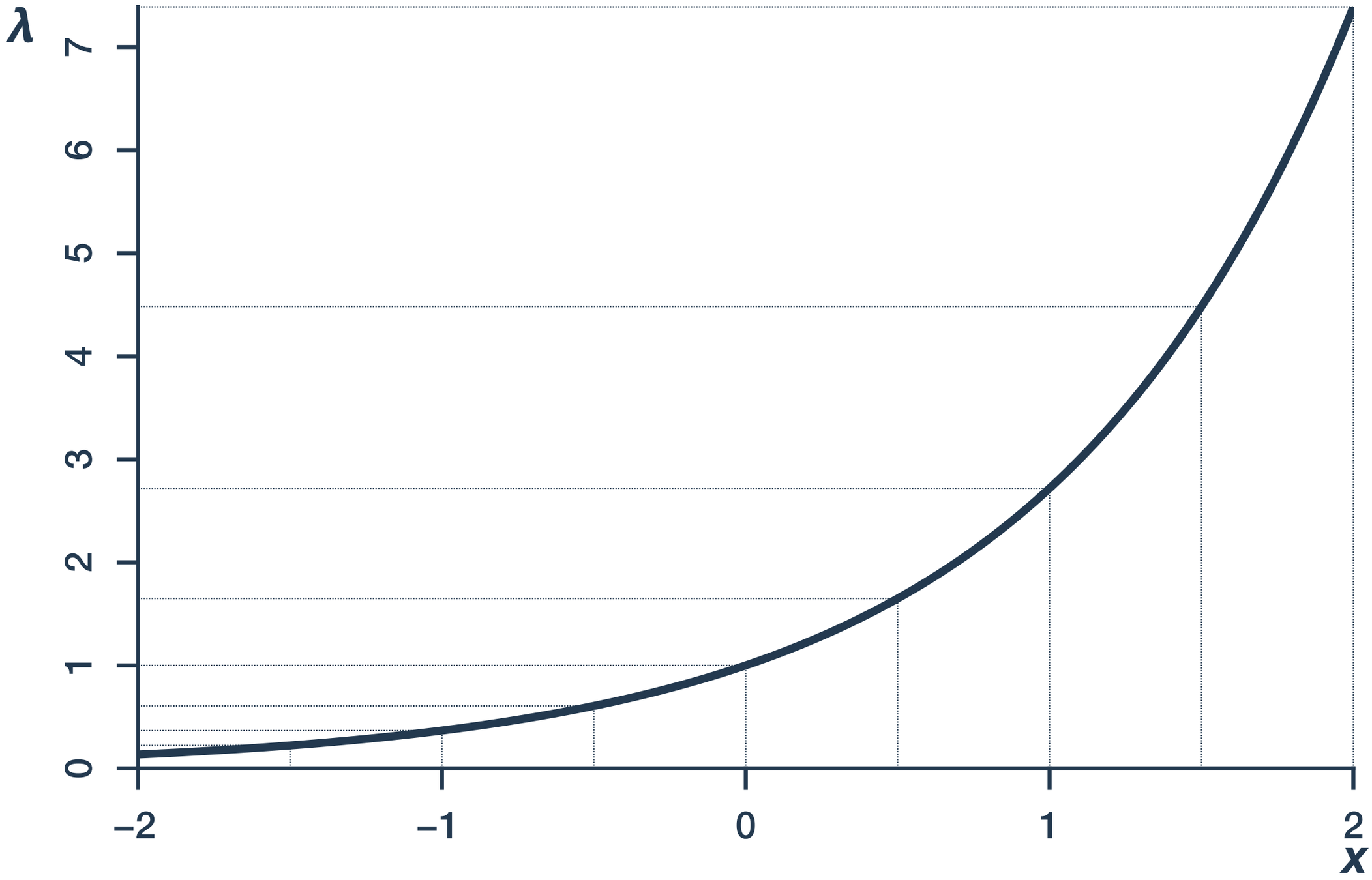
$$f(\lambda) = \alpha + \beta M_i$$

Log link function

$$\log(\lambda) = x$$

$$\lambda = e^{x}$$

$\lambda \in [0, \infty)$

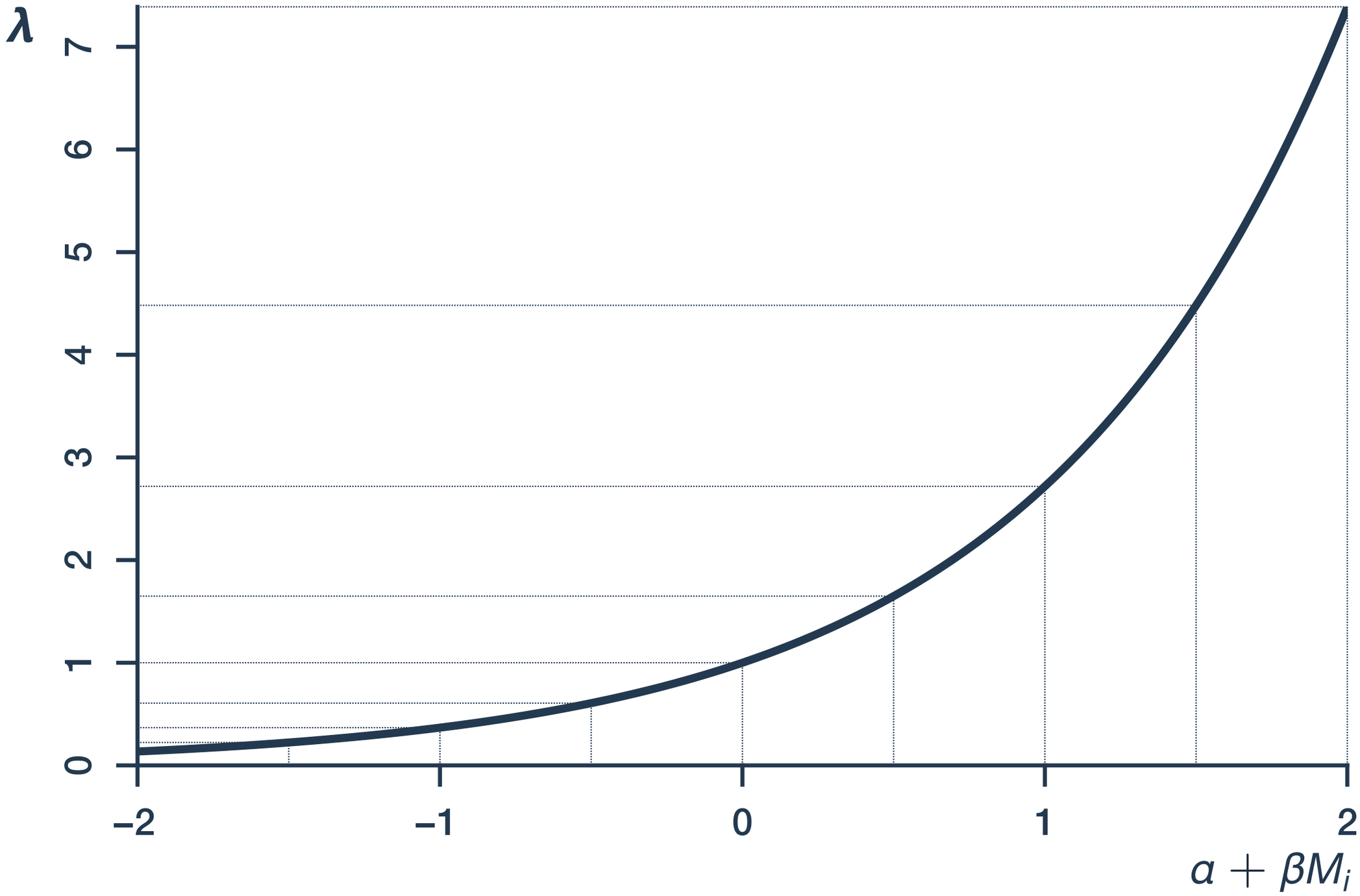


$x \in (-\infty, \infty)$

Log link function

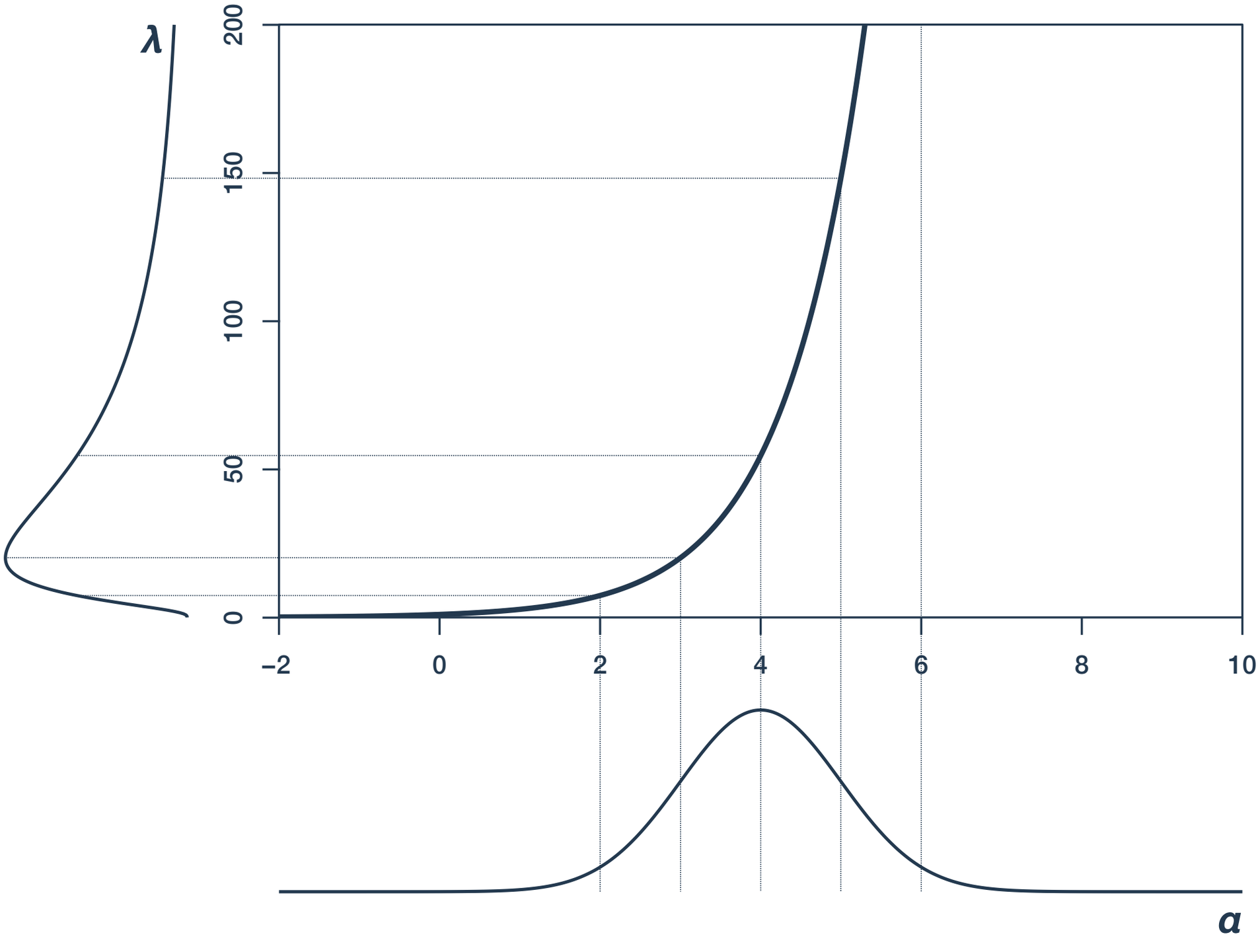
$$H_i \sim \text{Pois}(\lambda)$$

$$\log(\lambda) = \alpha + \beta M_i$$

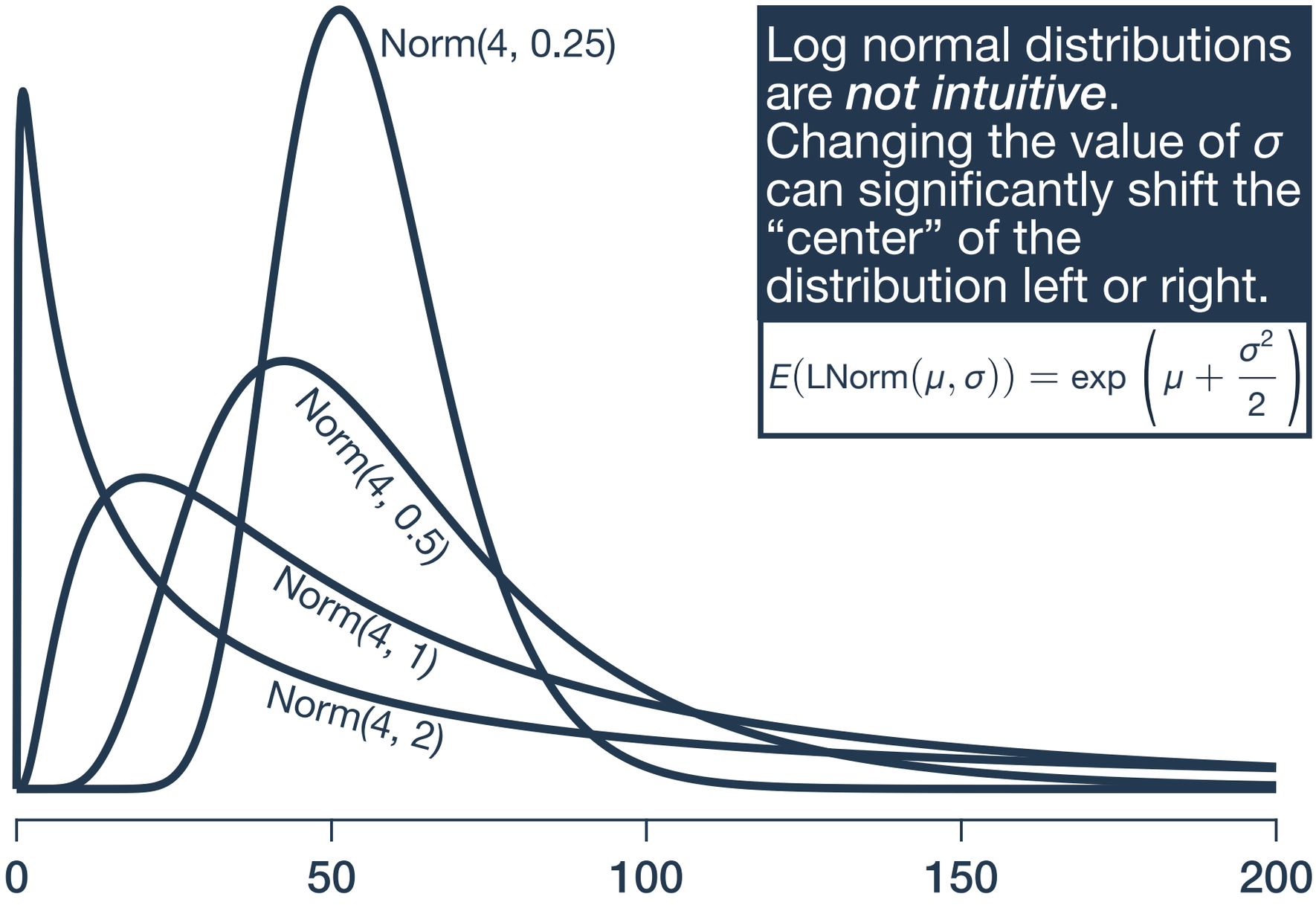


Log link priors

$$a \sim \text{Norm}(4, 1)$$



Log link priors

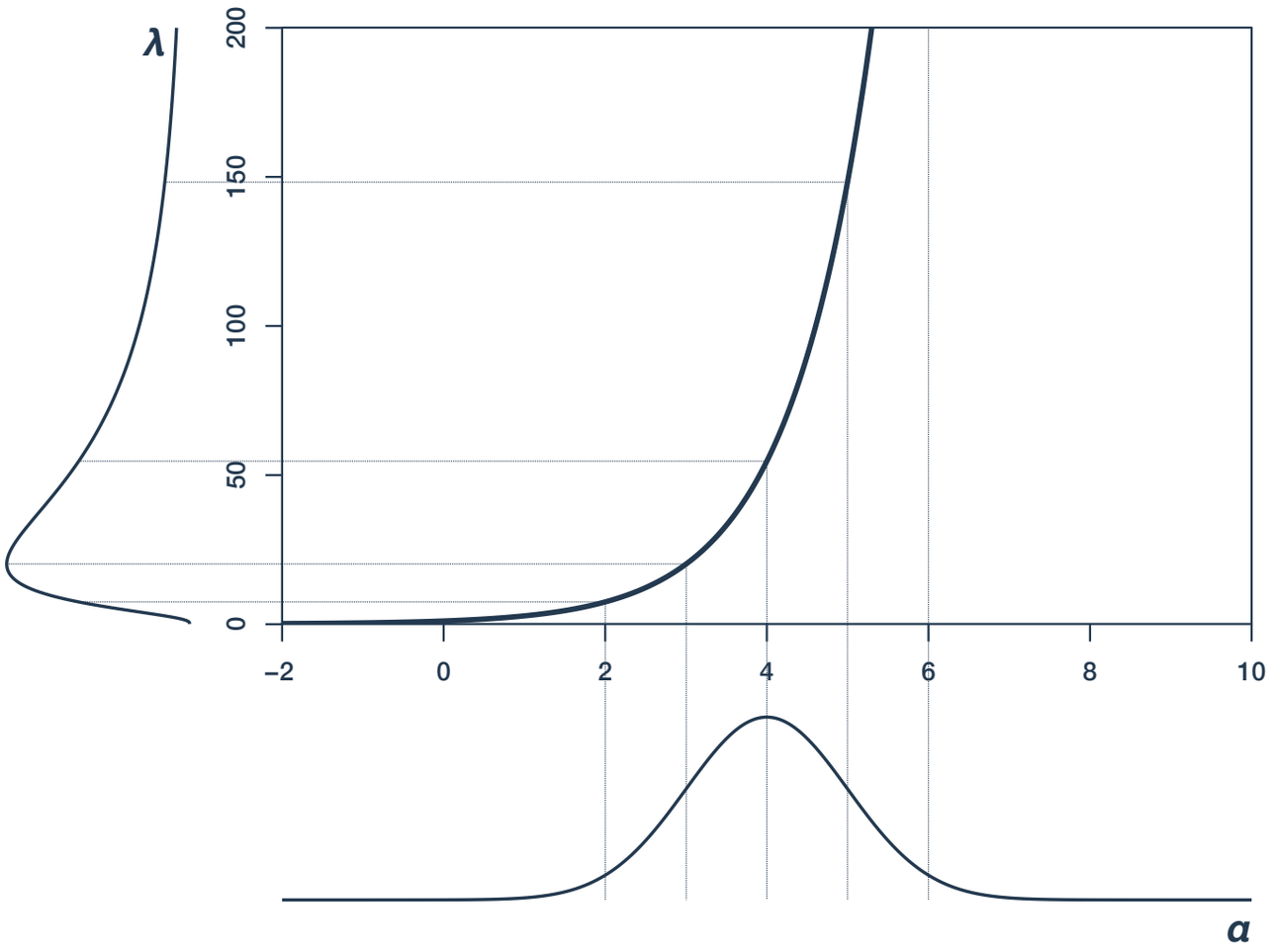


Log normal distributions are ***not intuitive***. Changing the value of σ can significantly shift the "center" of the distribution left or right.

$$E(\text{LNorm}(\mu, \sigma)) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

Log link priors

$$a \sim \text{Norm}(4, 1)$$



Compresses lefthand side of prior | Log link function maps $a \leq 0$ into the interval $\lambda \in [0,1]$

Expands righthand side of prior | Log link function maps $a \geq 6$ into the interval $\lambda \in [400,\infty)$

Prior depends on what is reasonable in data | Events you expect to be rare should use a narrower prior