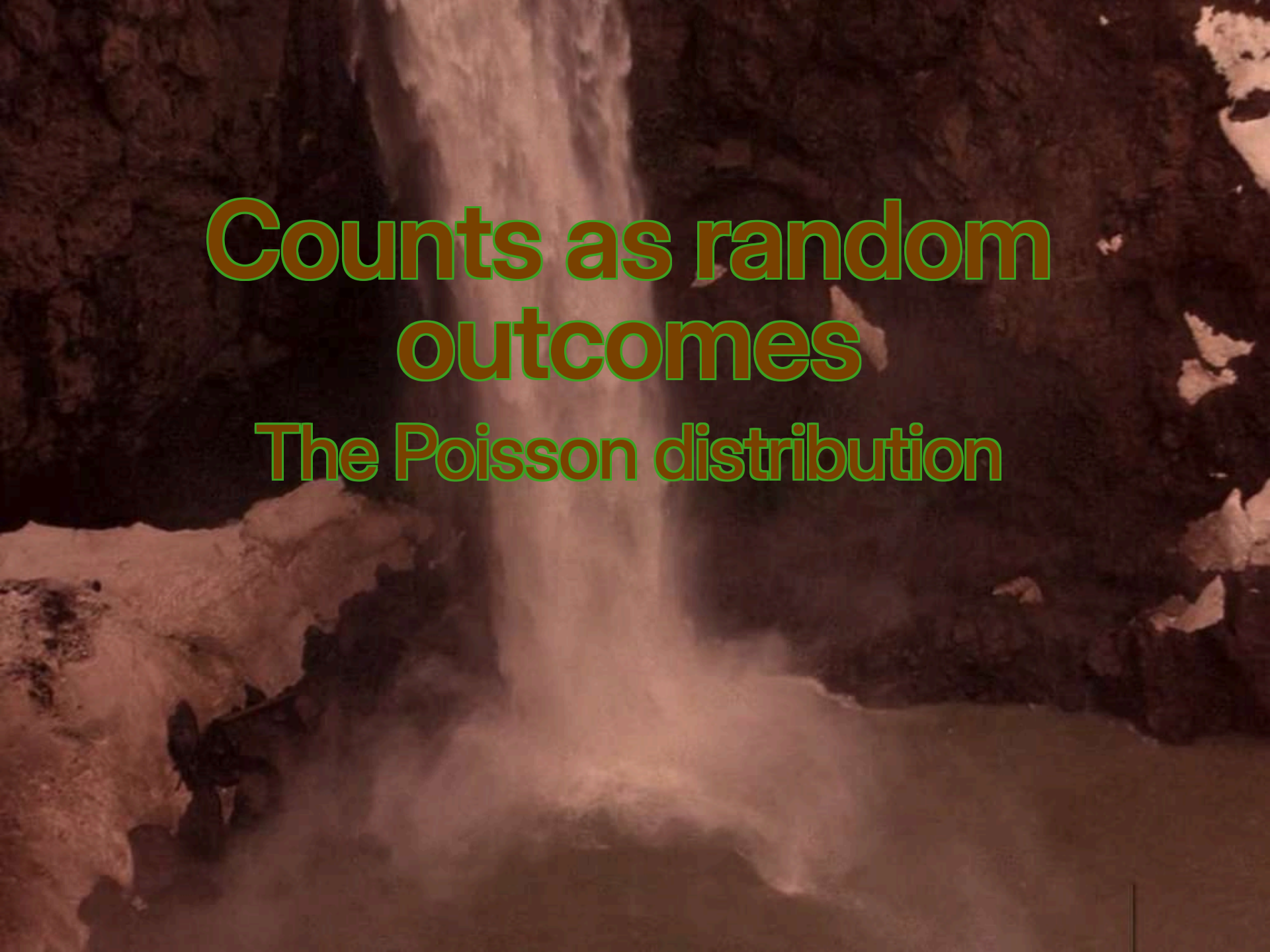
AGENDA

Counts & rates

1. Counts as random outcomes
2. The log link function
3. ***Hands on:*** *Poisson models in R*

Slides are licensed under CC BY-NC-SA 4.0
Uploading to generative AI tools or use for training machine learning models is prohibited



Counts as random outcomes

The Poisson distribution

Kinds of counts:

Average and deviation

An event that is technically a count, but the *scale* of the 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 occurence

An event that has no (theoretical) upper limit, but tends to happen at a relatively low rate

E.g. individual fertility, number of friends, grocery stores in a neighborhood, etc.

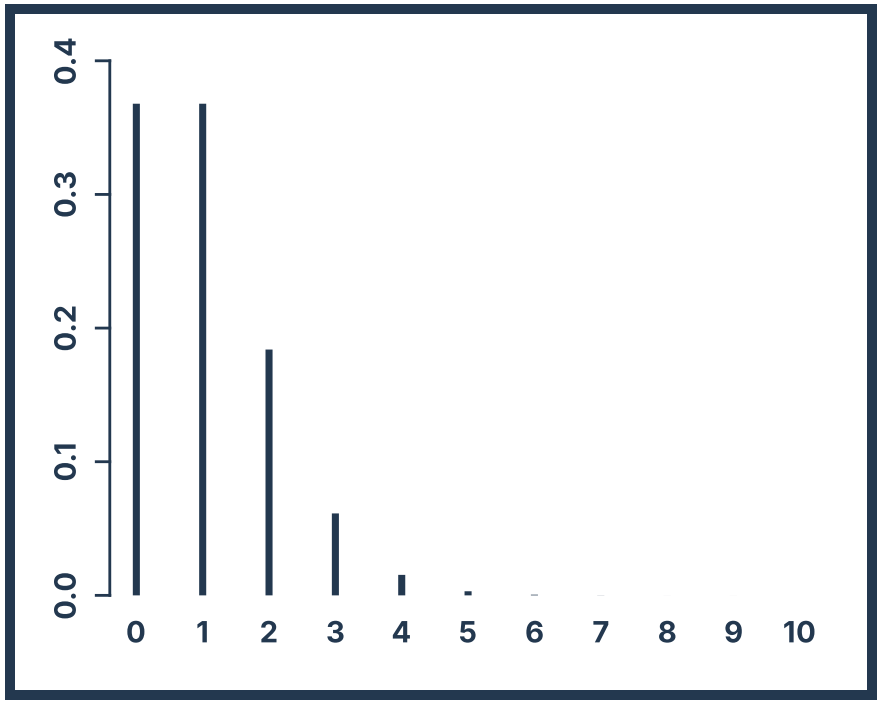
*(**See also**: geometric distribution for 'time until occurrence' types of models)*

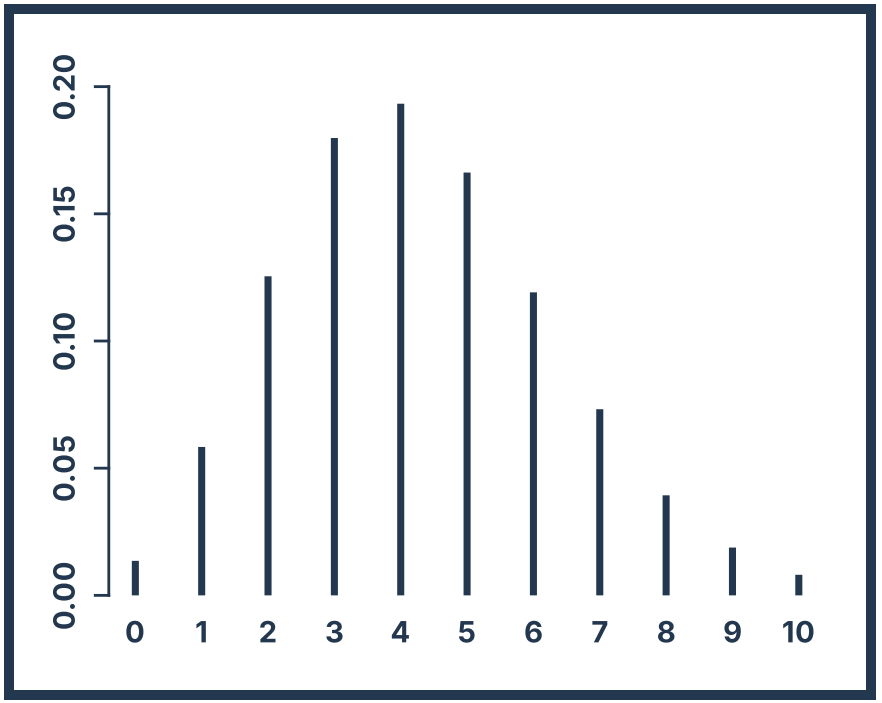
Poisson distribution

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.

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

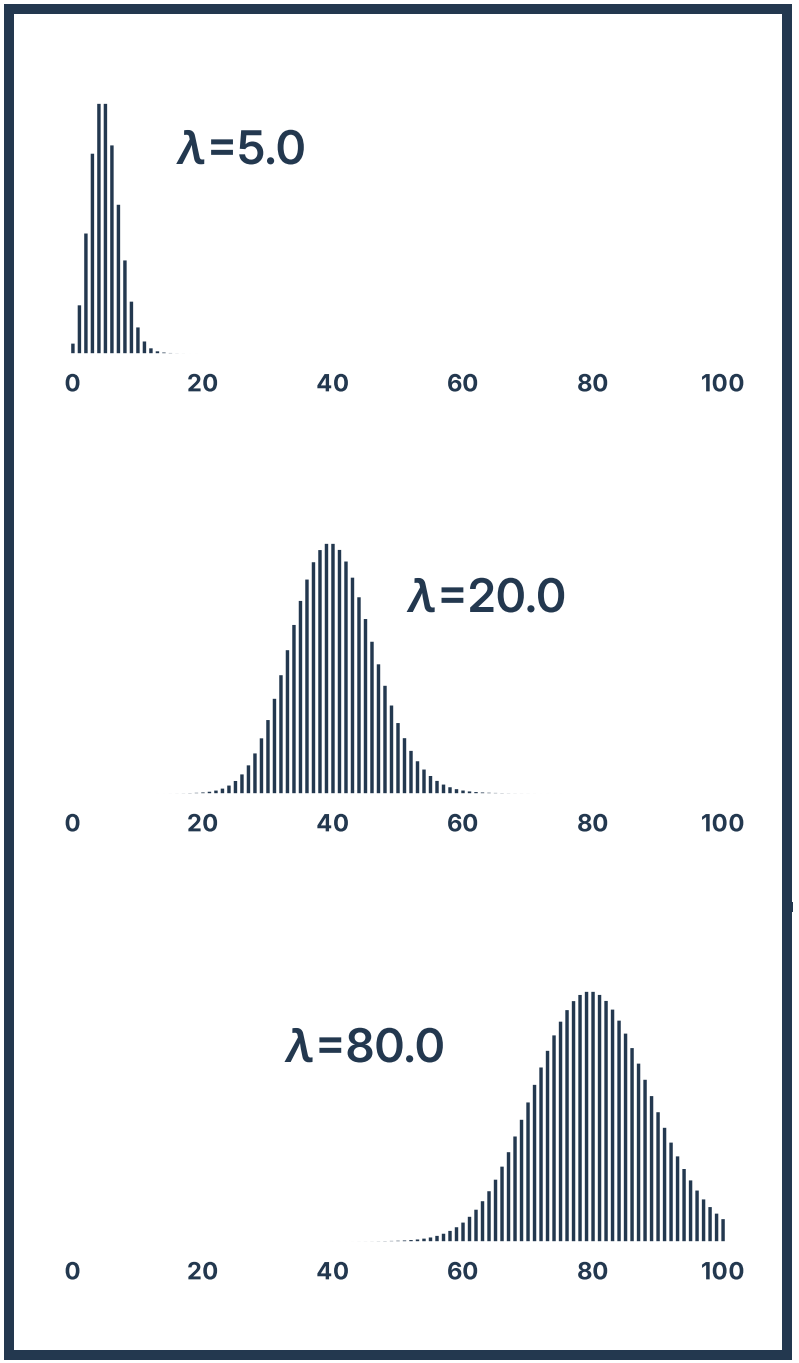




POISSON DISTRIBUTION

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

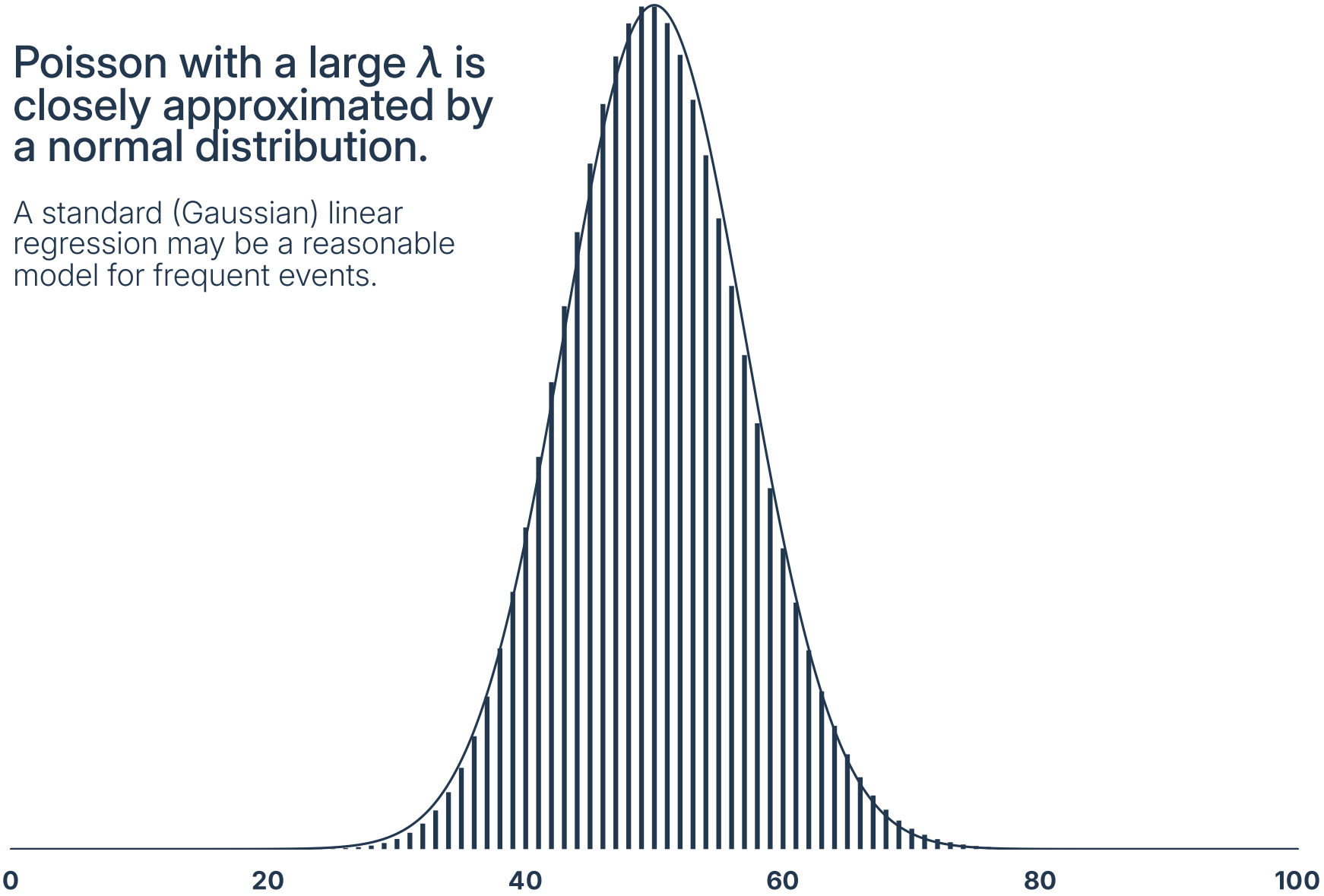
Single parameter:
λ is both the *mean* and *variance* (s. d. squared) of the Poisson distribution



POISSON DISTRIBUTION

Poisson with a large λ is closely approximated by a normal distribution.

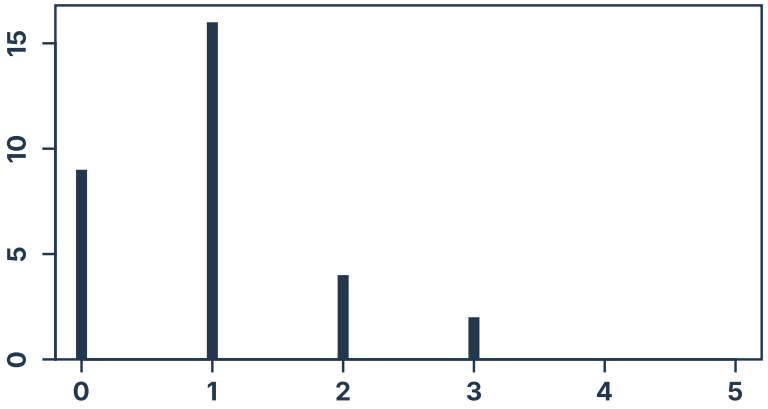
A standard (Gaussian) linear regression may be a reasonable model for frequent events.



POISSON-DISTRIBUTED OUTCOME

How many cups of coffee does Special Agent Dale Cooper drink in an episode of *Twin Peaks*?

(Seasons 1 and 2)

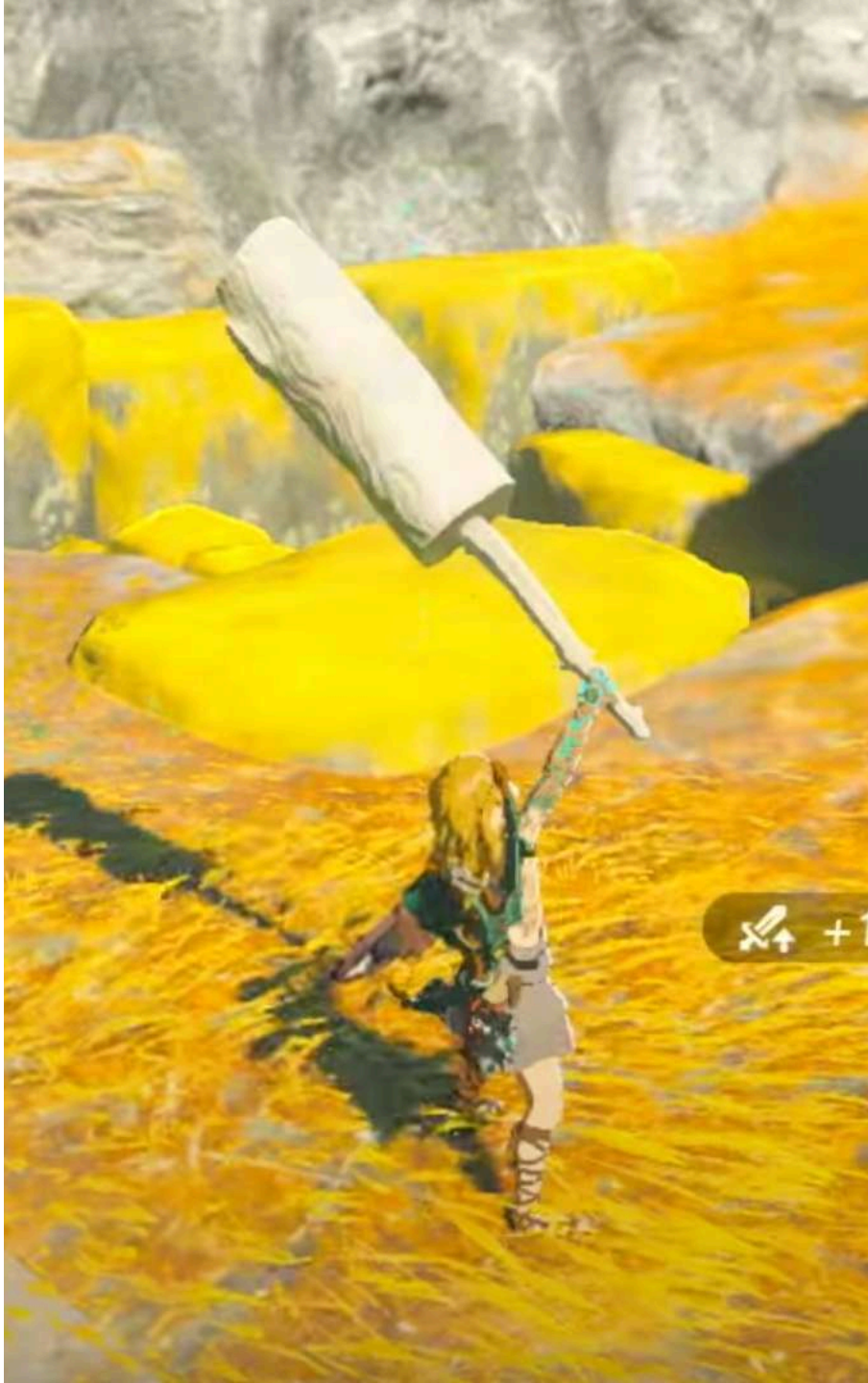


$$C_i \sim \text{Pois}(\lambda_i)$$

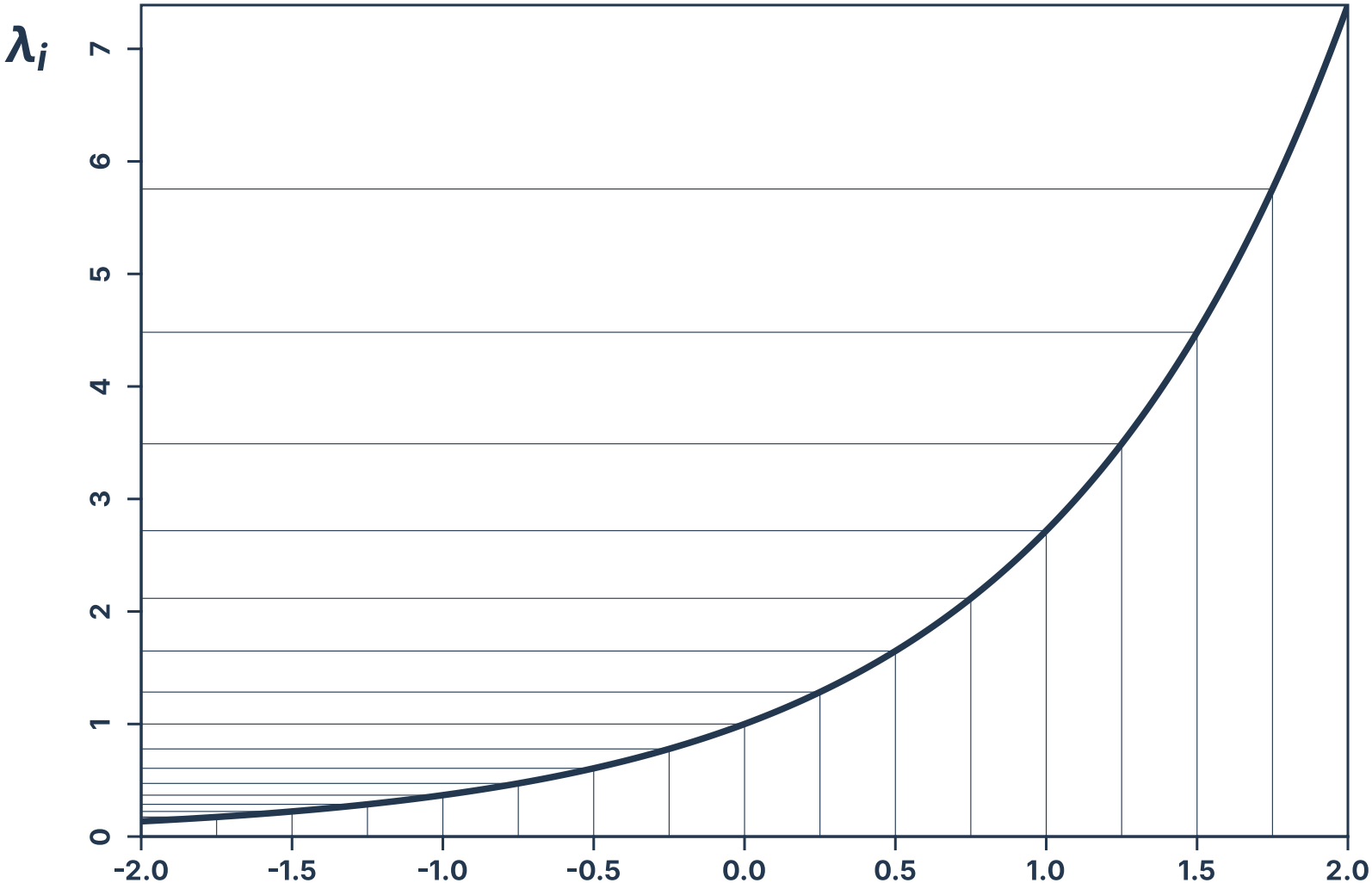
$$f(\lambda_i) = \alpha + \beta X_i + \dots$$

The log link

The log link



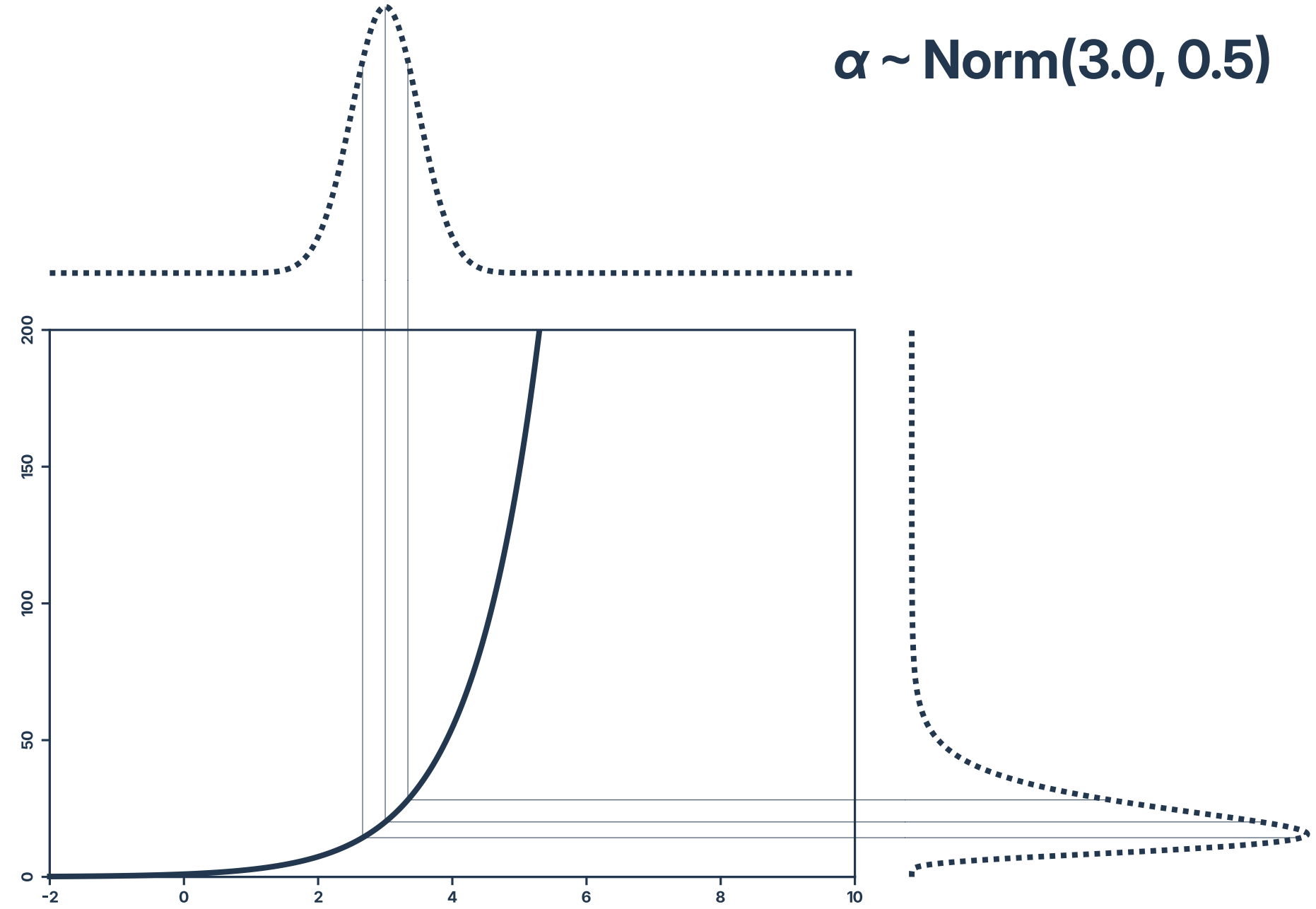
LOG LINK FUNCTION



$$\log(\lambda_i) = \alpha + \beta X_i + \ldots$$

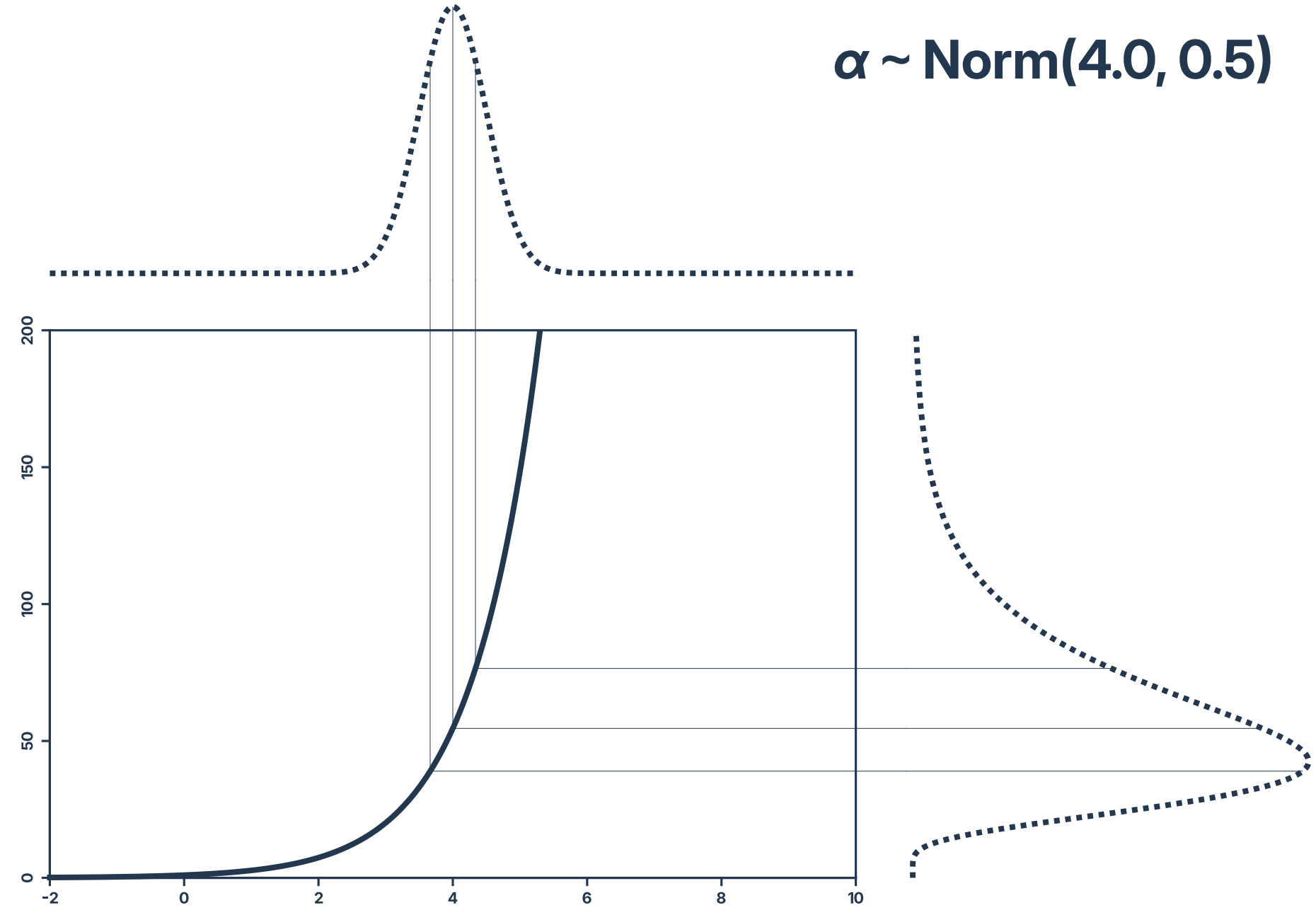
LOG LINK PRIORS

$\alpha \sim \text{Norm}(3.0, 0.5)$

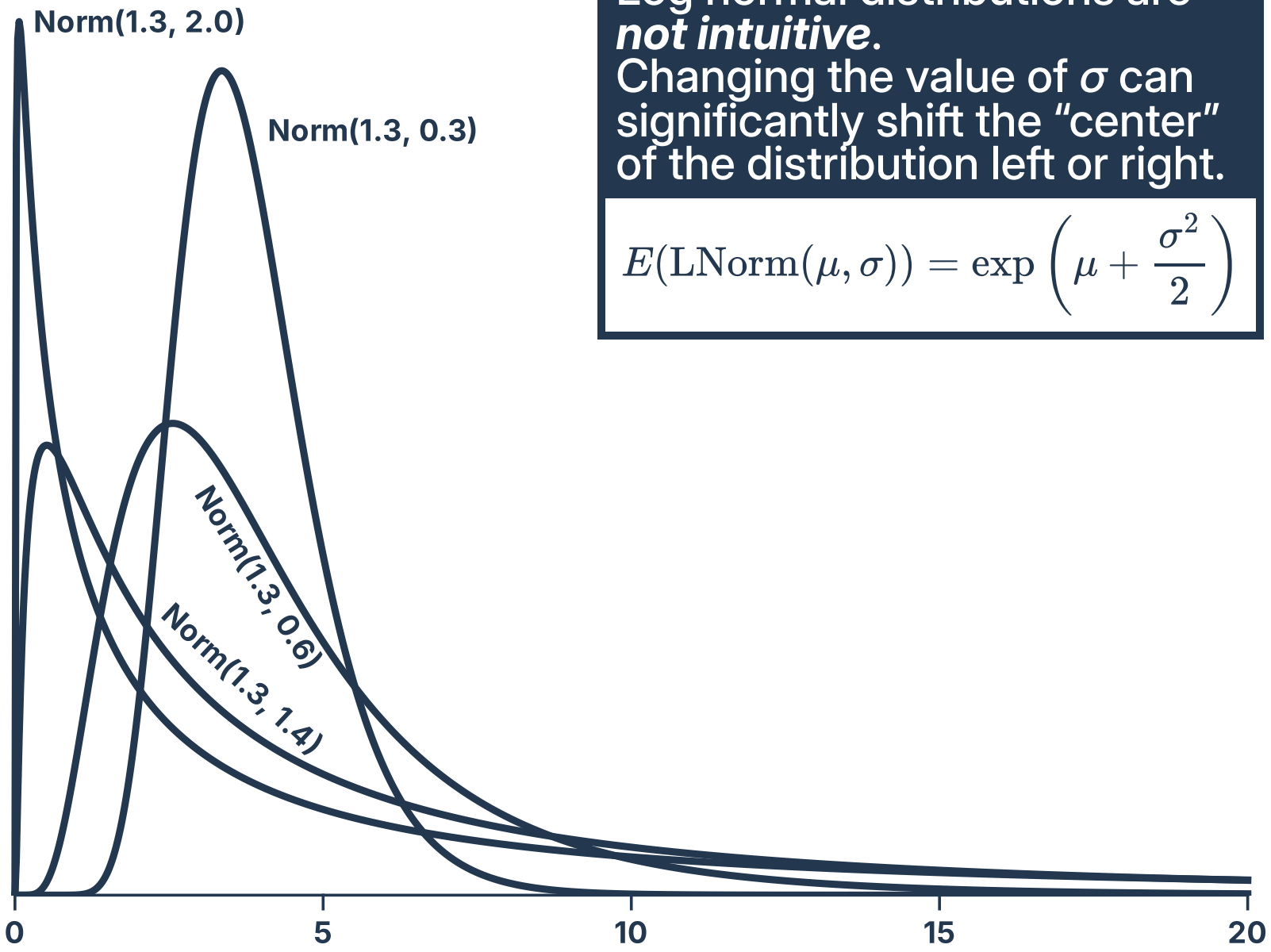


LOG LINK PRIORS

$\alpha \sim \text{Norm}(4.0, 0.5)$



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.

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

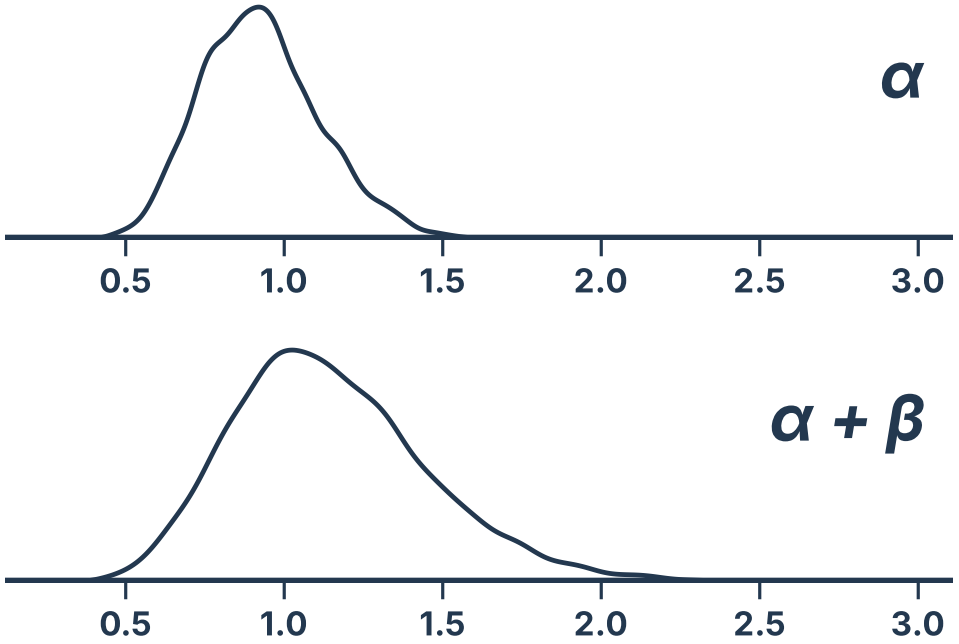
BACK TO THE MODEL

$$C_i \sim \text{Pois}(\lambda_i)$$
$$\log(\lambda_i) = \alpha + \beta S_i$$

$$\alpha \sim \text{Norm}(1.3, 1.4)$$
$$\beta \sim \text{Norm}(0, 0.5)$$

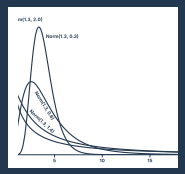
S_i – indicator for season 1

	Mean	95% C.I.
α	-0.097	(-0.513, 0.291)
β	0.200	(-0.414, 0.809)
$\exp(\alpha)$	0.908	(0.599, 1.337)
$\exp(\beta)$	1.221	(0.661, 2.247)



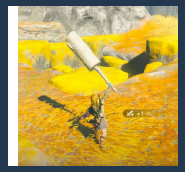
ACTION-PACKED EXCITEMENT FOR THE ENTIRE FAMILY
RECOMMENDED FOR AGES 8 TO ADULT.
NOW THE ENTIRE FAMILY CAN GET IN ON THE ACTION.
The Nintendo Entertainment System Action Set is a video system designed with the entire family in mind. That's because only Nintendo offers innovative accessories like the NES Advantage and NES Max for added excitement and a vast library of Game Paks that are both simple and sophisticated enough to challenge the abilities of everyone in the family.
Nintendo
ENTERTAINMENT SYSTEM®

Image credit

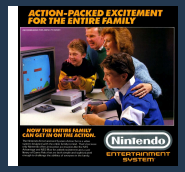


Figures by Peter McMahan (source code)

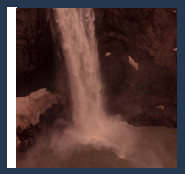
Still from Twin Peaks (1990)



Screenshot from The Legend of Zelda: Tears of the Kingdom (2023)



Nintendo ad via Reddit



Still from Twin Peaks (1990) via The beauty of TWIN PEAKS

Promotional image for Twin Peaks (1990)

Still from Twin Peaks (2017)