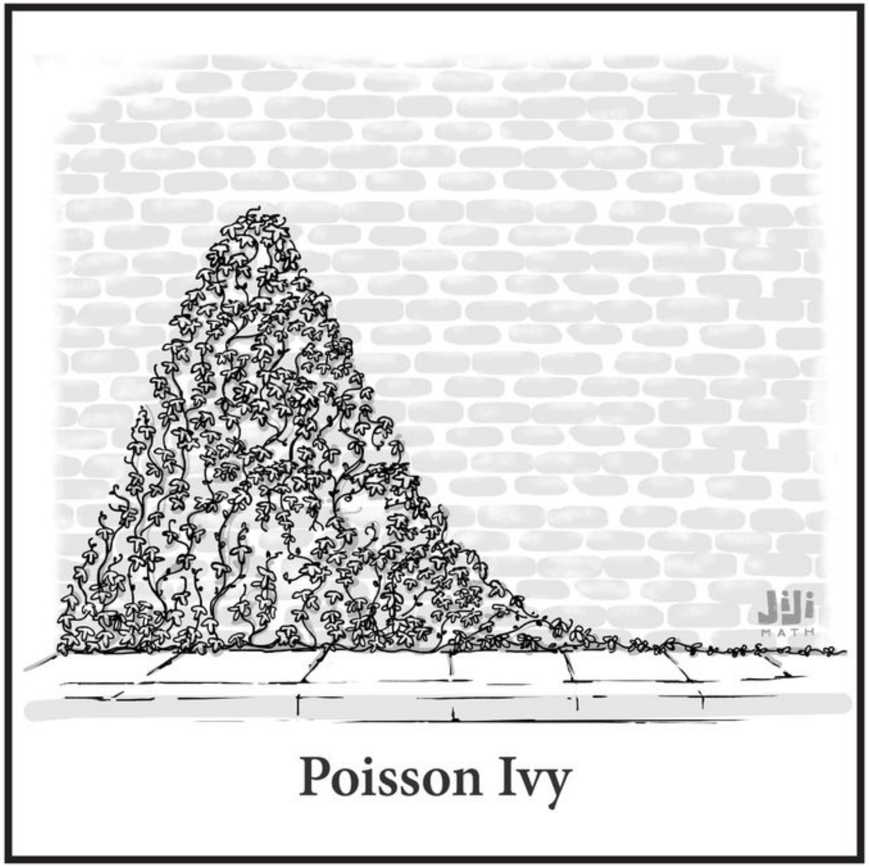
BST 210
Applied Regression Analysis



Lecture 20
Plan for Today

- What is a generalized linear model?
  - Linear, logistic, Poisson regression as special cases
  - The exponential family of distributions

- Poisson data
  - Poisson distribution
  - Poisson regression

New and Old: Generalized Linear Models

- Let's take some *very familiar* models and distributions, and parameterize/present them in a slightly more generalized way

- You will see that all models we have worked with thus far, as well as many of those upcoming, will then fall into a general unified framework

- This generalized framework enjoys great mathematical properties, which in turn facilitates inference on a broad scale

- Let's then delve briefly into a framework, or class, of models that we refer to as ***<u>Generalized Linear Models</u>*** →

Motivating Generalized Linear Models

- You'll see throughout, that the 3 main ingredients of a *Generalized Linear Model (GLM)* are:

  - A probability distribution for Y_i

  - A linear predictor $\mathbf{X}^T \boldsymbol{\beta} = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k = \eta$

  - A link function $g(\cdot)$

- Don't get distracted by the notation (which is required to describe properly), and keep mind on the broader picture

Motivating Generalized Linear Models

- Subtle (but critical) points about these 3 main ingredients to keep in mind as we go:

  - A probability distribution for Y_i

    (Best properties if from <u>Exponential Family of</u> <u>distributions</u>… most we deal with are…)

  - A linear predictor $\mathbf{X}^T \boldsymbol{\beta} = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k = \eta$

    (<u>Same</u> 'mean function' as before…)

  - A link function $g(\cdot)$

    (Ex: logit; Best properties if <u>canonical link</u>…and we usually restrict to be smooth (differentiable) and monotonic (invertible))

Motivating Generalized Linear Models

Why the canonical link? A canonical link will be one that

- Ensures that the linear predictor $\mathbf{X}^{T} \boldsymbol{\beta}$ is modeling the parameter of interest in the distribution

- Simplifies the derivation of the MLE

- Ensures that many great properties of the simplest case of the GLM, the linear model, hold (such as all errors summing to 0)

- Ensures that μ stays within the range of the outcome variable (recall what logit does for binary data)

We thus prefer this link when specifying a model.

Generalized Linear Model (GLM)

- A generalized linear model (GLM) consists of a linear predictor

$$\eta_i = \alpha + \sum_{k=1}^{K} \beta_k x_{ik},$$

a link function that describes how the mean, $E(Y_i) = \mu_i$, depends on the linear predictor,

$$\mathrm{g}(\mu_i) = \eta_i,$$

and a variance function,

$$Var(Y_i) = \phi V(\mu_i).$$

where $\phi > 0$ is a constant dispersion parameter.

Motivating context: Recall Linear Regression Model

- Assume Y_i is a continuous random variable, where

$$E(Y_i) = \alpha + \sum_{k=1}^{K} \beta_k x_{ik}, \text{ and}$$

$$Y_i = \alpha + \sum_{k=1}^{K} \beta_k x_{ik} + e_i, \text{ where}$$

$x_{ik} = k^{th}$ covariate for the i^{th} subject, and

$e_i =$ (independent) error term for the i^{th} subject

Motivating context: Recall Linear Regression Model

- We assume that the errors are independent and identically distributed, such that

$$E(e_i) = 0 \text{ and } Var(e_i) = \sigma^2.$$

Typically we also assume that

$$e_i \sim N(0, \sigma^2).$$

Linear Regression Model in Broader Context: as a GLM

- Linear regression uses the linear predictor

$$\eta_i = \alpha + \sum_{k=1}^{K} \beta_k x_{ik},$$

the identity function as the link function because

$$E(Y_i) = \mu_i = \eta_i, \text{ so}$$

$$g(\mu_i) = \mu_i = \eta_i$$

and $V(\mu_i) = 1$, yielding the variance function,

$$Var(Y_i) = \phi.$$

Logistic Regression Model in Broader Context: as a GLM

- Logistic regression uses the linear predictor

$$\eta_i = \alpha + \sum_{k=1}^{K} \beta_k x_{ik},$$

the logit function as the link function because

$$E(Y_i) = \mu_i = \exp(\eta_i)/(1+\exp(\eta_i)), \text{ so}$$

$$g(\mu_i) = \text{logit}(\mu_i) = \log(\mu_i/(1-\mu_i)) = \eta_i,$$

and $V(\mu_i) = \mu_i \times (1-\mu_i)$ with $\phi = 1$, yielding the variance function,

$$Var(Y_i) = \mu_i \times (1-\mu_i).$$

Poisson Regression Model (next!) as a GLM

- Poisson regression uses the linear predictor

$$\eta_i = \alpha + \sum_{k=1}^{K} \beta_k x_{ik},$$

the log function as the link function because

$$E(Y_i) = \mu_i = \exp(\eta_i), \text{ so}$$

$$g(\mu_i) = \log(\mu_i) = \eta_i,$$

and $V(\mu_i) = \mu_i$ with $\phi = 1$, yielding the variance function,

$$Var(Y_i) = \mu_i.$$

Allowing different exposure times is possible.

Quick look also at Probit Link

So that you've seen it…

- $\Phi^{-1}(p) = \theta$

  $p = \Phi(\theta)$

Where Φ denotes the cumulative distribution function of the standard normal density. For probit regression,

$$p = \mathrm{P}(Y = 1) = \Phi(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK})$$

What is this Exponential Family of Distributions?

- Many commonly used distributions, including the normal, Bernoulli, binomial, and Poisson, are members of the exponential family, whose densities can be written in the form

$$f(y;\theta,\phi)=\exp\left(\frac{y\theta-b(\theta)}{\phi+c(y,\phi)}\right)$$

where θ is the canonical parameter and ϕ is the dispersion parameter.

Exponential Family of Distributions

- Exponential family distributions can be either continuous (e.g., normal) or discrete (e.g., Bernoulli, binomial, Poisson).

  Other exponential family distributions include the chi square, exponential (survival curves), gamma, and beta distributions.

  Some distributions that are <u>not</u> exponential family include student's *t* and the uniform distributions.

Exponential Family of Distributions

- For exponential family distributions

$$f(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi + c(y,\phi)}\right),$$

and it can be shown that

$$E(Y) = b'(\theta) = \mu$$

and

$$Var(Y) = \phi\, b''(\theta) = \phi\, V(\mu).$$

Exponential Family of Distributions

- Above, θ depends on expected value of Y, E(Y), so is related to the 'mean'; whereas ϕ is a scale parameter, and related to the 'variance'; also b, c are arbitrary functions

- If we can '*massage*' a distribution function we are working with into the form of an *Exponential Family Distribution*, then we can be assured that our problem falls within the *Generalized Linear Model framework*

- Then maximum likelihood estimation and inference are greatly simplified and can be handled in this unified framework

- Also, the Log-Likelihood of Exponential Family distributions will always be concave—thus we are guaranteed one maximum—great result when it comes to parameter estimation!

Example: Normal(θ, ϕ) as Exponential Family

- $$f(y;\theta,\phi) = (2\pi\phi)^{-1/2}\exp\left(-\frac{(y-\theta)^2}{2\phi}\right)$$

$$= (2\pi\phi)^{-1/2}\exp\left(-\frac{y^2}{2\phi}\right)\exp\left(\frac{y\theta-\frac{1}{2}\theta^2}{\phi}\right)$$

$$= \exp\left(\frac{y\theta-\frac{1}{2}\theta^2}{\phi+c(y,\phi)}\right)$$

where $b(\theta)=\frac{1}{2}\theta^2$ and $c(y,\phi)$ is complicated

but does not involve θ.

Example: Normal(θ, ϕ) as Exponential Family

- For $b(\theta) = \dfrac{1}{2}\theta^2$ we have

$$E(Y) = b'(\theta) = \theta$$

and

$$Var(Y) = \phi\, b''(\theta) = \phi.$$

Example: Bernoulli(*p*) as Exponential Family

- Let $p = \exp(\theta)/((1+\exp(\theta))$. Then

$$f(y;\theta,\phi=1) = p^y\,(1-p)^{1-y} \text{ for } y = 0, 1$$

$$= (\exp(\theta))^y/(1+\exp(\theta))$$

$$= \exp\big(y\theta - \log(1+\exp(\theta))\big)$$

where $b(\theta) = \log(1+\exp(\theta))$ and $c(y,\phi) = 0$.

Then $E(Y) = b'(\theta) = \exp(\theta)/(1+\exp(\theta)) = p$ and

$Var(Y) = 1\times b''(\theta) = \exp(\theta)/(1+\exp(\theta))^2 = p(1-p)$.

Example: Poisson(λ) as Exponential Family

- Let $\lambda = \exp(\theta)$ and $\theta = \log(\lambda)$. Then

$$f(y;\theta,\phi=1) = \exp(-\lambda)\lambda^y / y! \text{ for } y = 0, 1, 2, \ldots$$
$$= \exp(-\lambda + y\log(\lambda) - \log(y!))$$
$$= \exp\left(y\theta - \exp(\theta) - \log(y!)\right)$$

where $b(\theta) = \exp(\theta) + \log(y!)$ and $c(y,\phi) = 0$.

Then $E(Y) = b'(\theta) = \exp(\theta) = \lambda$ and

$Var(Y) = 1 \times b''(\theta) = \exp(\theta) = \lambda$.

Advantages of GLMs

- Provides a flexible generalization of linear regression that allows the outcomes to have distributions other than normal

- Puts linear, logistic, Poisson, and other regression methods into a unified framework

- Estimation methods through maximum likelihood or iteratively reweighted least squares (sometimes method of moments for ϕ)

Advantages of GLMs

- Generalized linear models with canonical links have desirable statistical properties and hence tend to be used by default in most statistical packages

  – identity link for linear regression
  – logit link for logistic regression
  – log link for Poisson regression

- In particular, approximate normality of β coefficients on this scale tends to work well, hence CIs and *p*-values are based on this scale

Summary/Advantages of GLMs

- For binary outcomes, the logit link is the canonical link function

$$\text{logit}(p) = \log(p \,/\, (1 - p)) = \theta$$

$$p = \exp(\theta) \,/\, [1 + \exp(\theta)]$$

- But, other links might be useful in certain contexts, e.g., for binary outcomes we could use
  - probit link
  - complementary log-log link, or even
  - identity link
  - log link

Summary/Advantages of GLMs

- Other extensions come easily to GLMs, including
  - Models for “over dispersion” or heterogeneity of variances
  - Generalized estimating equations (GEEs) to accommodate correlations between responses, to get a “population-averaged” effect of covariates
  - Generalized linear mixed models (GLMMs) that allow random effects in the linear predictor

Coming Up Next

- Poisson data and regression!

Introduction to Poisson Regression

We will consider

- The Poisson probability distribution and the modeling of rates

- Two group comparisons and incidence rate ratios

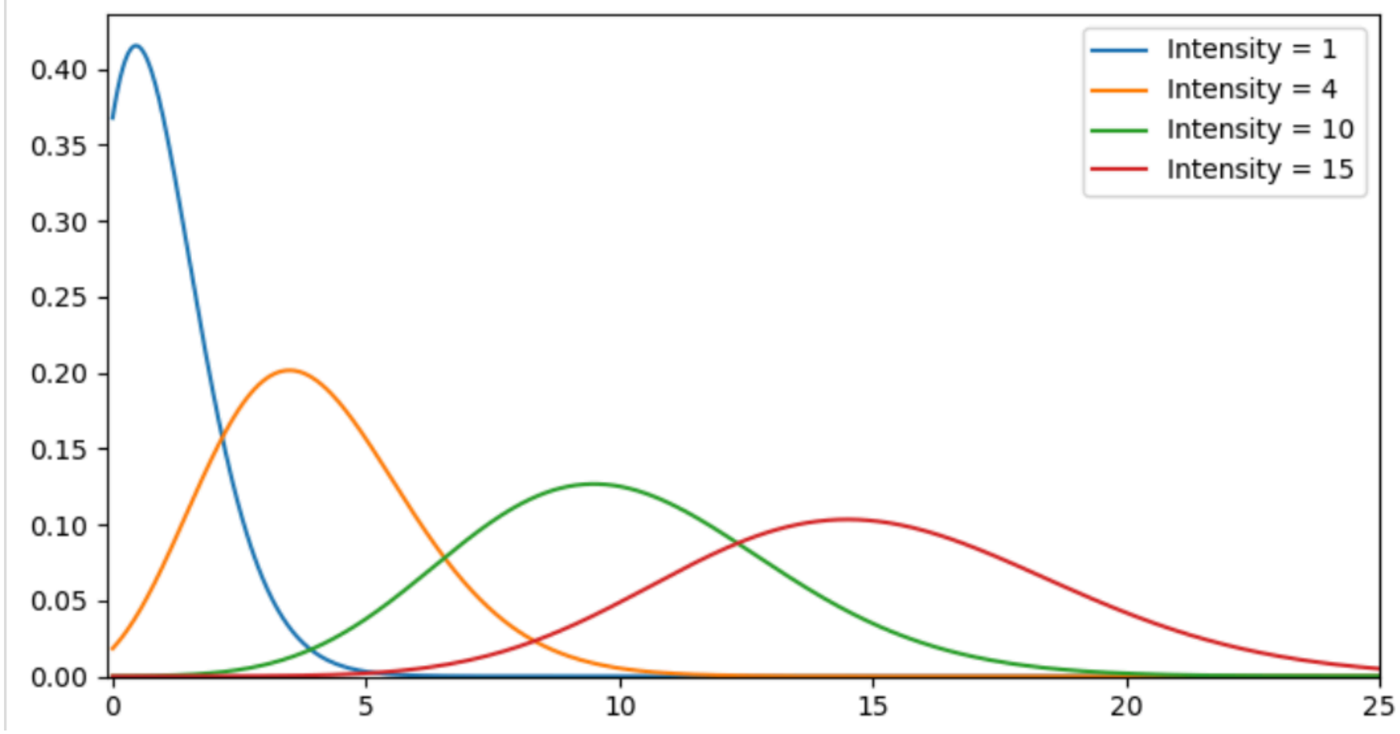
- Introduction to Poisson regression modeling

And even more after that: Zero-inflated Poisson regression, Negative binomial regression, Over dispersed Poisson…

Data arising from Poisson Processes

- Number of births expected in 24 hours at a hospital
- Number of cases of disease in a certain town
- Number of mutations in given regions of a chromosome
- Number of particles emitted by radioactive source in a given time
- Number of patients with relapse within 1st year of treatment

Quick visual: Poisson Distribution



Motivating Example: Rare Cancer

- Suppose a hospital has observed 10 cases of a certain rare cancer over the past 20 years (has, on average, λ = 0.5 cases per year)
- It suddenly observes 3 cases in 1 year
- What is the probability of this occurrence?
- The hospital serves a large population; we do not know the exact number of people (but we assume the population size is constant over time)

Poisson Distribution

- We are interested in modeling the distribution of the number of new cases of cancer (Y) over some long time period (t) (e.g., 1 year)

- Let's assume

  $P(1 \text{ new case in time } \Delta t) \approx \lambda\, \Delta t$

  where Δt is a short time interval (e.g., 1 day, or even 1 minute or 1 second)

- λ = number of cases per unit time = incidence rate (just like in Epi!)

Poisson Distribution

- That is,

  $P(1 \text{ new case in time } \Delta t) \approx \lambda\, \Delta t$

  $P(0 \text{ new cases in time } \Delta t) \approx 1 - \lambda\, \Delta t$

  $P(2 \text{ or more new cases in time } \Delta t) \approx 0$

- We'll also assume

  Stationarity: Incidence rate stays the same over time (λ is a constant)

  Independence: (where Δt_1 and Δt_2 do not overlap)

  $P(1 \text{ new case in } \Delta t_1 \ \underline{\text{and}}\ 1 \text{ new case in } \Delta t_2)$

  $= P(1 \text{ new case in } \Delta t_1) \times P(1 \text{ new case in } \Delta t_2)$

Poisson Distribution

- Based on these assumptions, it can be shown that

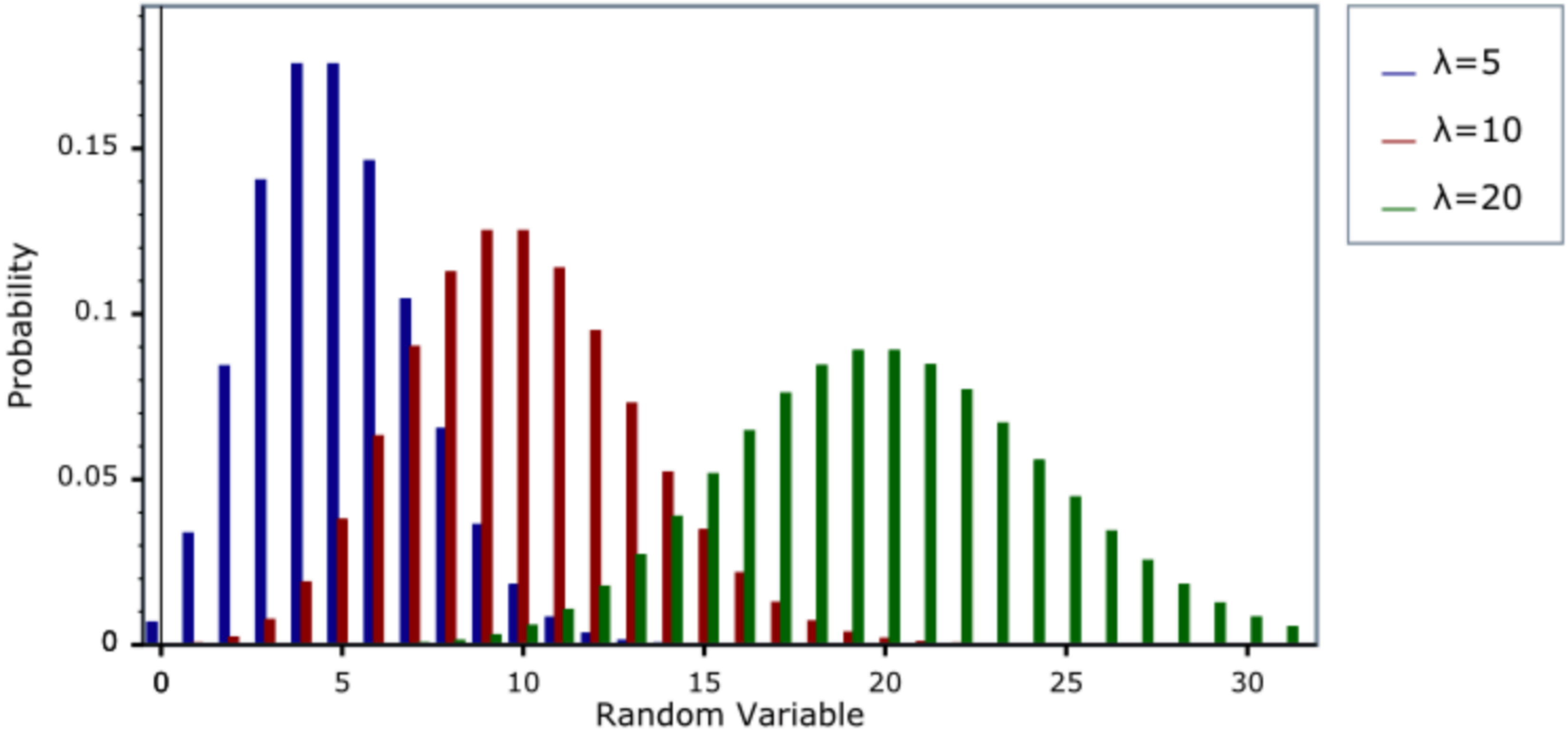
$$P(Y = y) = \frac{e^{-\lambda t}(\lambda t)^y}{y!} \qquad \text{where } y = 0, 1, 2, \ldots$$

- Note that this distribution is a function of $\mu = \lambda t$ where
  μ = expected number of cases over time t
  = number of cases per unit time $\times$ time interval

- Thus we can also write

$$P(Y = y) = \frac{e^{-\mu}\mu^y}{y!} \qquad \text{where } y = 0, 1, 2, \ldots$$

Poisson Distribution



Motivating Example: Rare Cancer

- We can then model the number of new cases of rare cancer in 1 year using a *Poisson distribution* with $\lambda = 0.5$ and $t = 1$, or $\mu = (0.5)(1) = 0.5$

- $P(Y = 0) = e^{-0.5} (0.5)^0 / 0! = 0.607$

  $P(Y = 1) = e^{-0.5} (0.5)^1 / 1! = 0.303$

  $P(Y = 2) = e^{-0.5} (0.5)^2 / 2! = 0.076$

Motivating Example: Rare Cancer

- The question we are really asking is, "Is it unusual to observe 3 or more cases of cancer in any 1 year?"

- Thus we consider

$$
\begin{aligned}
P(Y \geq 3) &= 1 - P(Y \leq 2) \\
&= 1 - [P(Y = 0) + P(Y = 1) + P(Y = 2)] \\
&= 1 - 0.986 \\
&= 0.014
\end{aligned}
$$

- By convention, we might say that an event that happens less than 5% of the time is unusual

Recall from GLMs: Exponential Family of Distributions

- Many commonly used distributions, including the normal, Bernoulli, binomial, and Poisson, are members of the exponential family, whose densities can be written in the form

$$f(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi + c(y,\phi)}\right)$$

where θ is the canonical parameter and ϕ is the dispersion parameter.

Recall from GLMs: Exponential Family of Distributions

- For exponential family distributions

$$f\,(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi + c(y,\phi)}\right),$$

and it can be shown that

$$E(Y) = b'(\theta) = \mu$$

and

$$Var(Y) = \phi\, b''(\theta) = \phi\, V(\mu).$$

Recall Poisson (μ) is in Exponential Family

- We can show this by letting $\mu = \exp(\theta)$ and so $\theta = \log(\mu)$,
- then, for $y = 0, 1, 2, \ldots$ we have

$$
\begin{aligned}
P(Y = y) &= e^{-\mu} \mu^{y} / y! \\
&= e^{-\exp(\theta)} (\exp(\theta))^{y} / y! \\
&= \exp(y\theta - \exp(\theta) - \log(y!)) \\
&= \exp((y\theta - b(\theta)) / (\phi + c(y,\phi))
\end{aligned}
$$

where $b(\theta) = \exp(\theta) + \log(y!)$, $\phi = 1$, and $c(y,\phi) = 1$.

Poisson Expected Value and Variance

- Using exponential families properties, we find that

$$E(Y) = b'(\theta) = \exp(\theta) = \mu$$

and

$$\mathrm{Var}(Y) = \phi \times b''(\theta) = 1 \times \exp(\theta) = \mu$$

- Thus for the Poisson distribution (crucial result!)

$$E(Y) = \mu = \lambda\, T$$

and

$$\mathrm{Var}(Y) = \mu = \lambda\, T$$

Back to Example: Rare Cancer

- If the incidence rate per year is $\lambda = 0.5$, how many new cases of cancer would we expect to see over a 10-year period?

  $$E(Y) = 0.5(10) = 5$$

- The variance of Y is also 5, and the standard deviation is $\sqrt{5}$, or 2.236

- Now, how would we think about incorporating more than simply the outcome data? What about covariates? →

Extending to Poisson Regression Modeling

- That is, thinking forward, what ways can we model counts like this in a regression model?

- We already know that the Poisson distribution is in the exponential family, so we can employ the GLM framework.

- What link function makes the most sense?

- Would our old friend $E(Y) = \beta_0 + \beta_1 x_i + \dots$ work? Why or why not?

Extending to Poisson Regression Modeling

- If we use the canonical link $g(\cdot) = \log(Y)$, our transformed outcome to model is then

$$\log(E[Y]) = \log(\lambda t) = \log(\lambda) + \log(t)$$

- Since t is fixed/constant, it is not a function of our predictors. Thus the expected number of events E[Y] depends on our covariates of interest only through the incidence rate λ,

$$\log(\lambda) = \beta_0 + \beta_1 x_i + \ldots + \beta_p x_p$$

- Combining the above gives our Poisson regression model

$$\log(E[Y]) = \beta_0 + \beta_1 x_i + \ldots + \beta_p x_p + \log(t)$$

Extending to Poisson Regression Modeling

- Fitting this Poisson regression model

$$\log(E[Y]) = \beta_0 + \beta_1 x_i + \ldots + \beta_p x_p + \log(t)$$

we make the assumptions:

- The observed counts Y are independent
- Y ~ Poisson(λt), so the mean = variance
- The mean model is correctly specified
- The incidence rate λ is time-invariant (does not depend on the total person-time of exposure)

- Also log(t) is the 'offset' →

Extending to Poisson Regression Modeling

- Note that log(t), the 'offset':

- Adjusts for the fact that the total number of observed events depends both on the underlying incidence rate (the rate/number of events per some unit amount of time) and the total person-time of exposure

- Each individual or covariate pattern has a different amount of exposure time—by modeling the expected number of events in each individual/covariate pattern, it's important that we take this differing exposure time into account. The offset does this!

Next: Poisson Approximation to Binomial Distribution

- If $n \geq 100$ and $p < 0.01$ (this is a "rule of thumb"), we can approximate a binomial distribution with parameters n and p by a Poisson distribution with parameter $\mu = np$

- This works because the variance of a binomial random variable npq (ie np(1-p)) is approximately equal to the mean np if p is small and q is close to 1 (where $q = 1 - p$)

Poisson Approximation to Binomial Distribution

- Suppose that a hospital serves a population of exactly n = 25,000 people
- And the probability that any one of them is diagnosed with a certain rare cancer over the next year is $p = 3.6 \times 10^{-5}$
- Assuming a binomial distribution, $E(Y) = np = 0.9$
- To calculate the probability of observing at least 5 cases of cancer in 1 year, we could instead use the Poisson distribution with $\mu = np = 0.9$

Poisson Approximation to Binomial Distribution

- For X ~ Binomial(n = 25,000, p = 3.6 x 10^{-5}) then

  $P(X \geq 5) = 0.0023435$

- For X ~ Poisson(μ = 0.9) then

  $P(X \geq 5) = 0.0023441$

- These probabilities are very similar, and both are less than 0.05; therefore, observing 5 events in 1 year is unusual

- Note: Given computers, this approximation may not really be needed, but could be helpful in some situations

Normal Approximation to the Poisson

- Suppose there is a rare form of influenza which yields 15 cases per year on average in a certain community

- In 2017 there are 25 cases

- Is this an unusual occurrence?

- We could use the Poisson distribution to answer this question

- Y is a random variable representing the number of events in 1 year

Normal Approximation to the Poisson

- Note that $\lambda = 15$ cases per year and $T = 1$ year

- Therefore, $\mu = 15$

- We want to compute

$$P(Y \geq 25 \mid \mu = 15) = 1 - P(Y \leq 24)$$

$$\text{vv}= 1 - \sum_{k=0}^{24} e^{-15} (15)^k / k!$$

- We can also use a normal approximation to the Poisson distribution

Normal Approximation to the Poisson

- A Poisson random variable Y with parameters λ and T has mean $\mu = \lambda T$ and variance $\sigma^2 = \lambda T = \mu$

- We can approximate the Poisson Y by a normal random variable W with mean = variance = μ

- We can approximate $P(Y = k)$ by

  $P(k - 0.5 \leq W \leq k + 0.5)$ using a *continuity correction*

- This approximation should be used only if $\mu \geq 10$ (a "rule of thumb")

Example: Influenza

- We approximate a Poisson random variable Y with parameter $\mu = 15$ by a normal random variable W with mean 15 and variance 15

- Thus P($Y \geq 25$) is approximated by

$$
\begin{aligned}
P(W \geq 24.5) \quad &= P(Z \geq [24.5 - 15] / \sqrt{15}) \\
&= P(Z \geq 2.45) \\
&= 0.00714281
\end{aligned}
$$

- It is unusual to observe 25 cases in one year; there appears to be a significant increase in the number of cases in 2017

Example: Influenza

- Compare this with the exact Poisson calculations when Y ~ Poisson(15). Here, $P(Y \geq 25) = 0.01116478$.

- This approximation is accurate to the second decimal place, and would be more accurate with larger μ (here μ is only 15)

- Note: Given computers, this approximation may not really be needed, but could be helpful in some situations