Precept 6: Models for Duration and Count Data

Soc 504: Advanced Social Statistics

Ian Lundberg

Princeton University

March 16, 2017

Outline

Outline

Suppose you want to model the time until an event occurs.

Suppose you want to model the time until an event occurs.

Time has to be positive

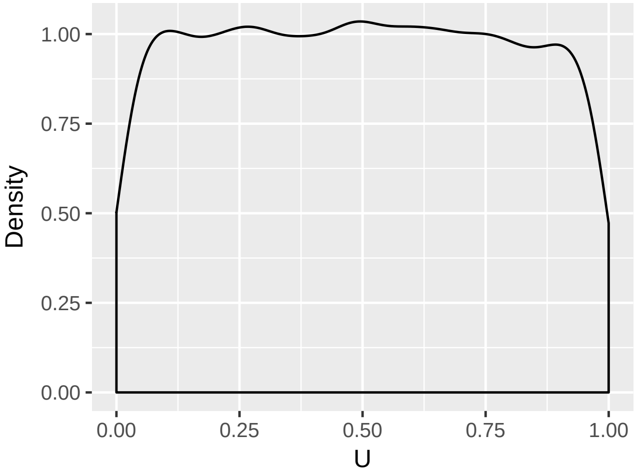
Suppose you want to model the time until an event occurs.

Time has to be positive

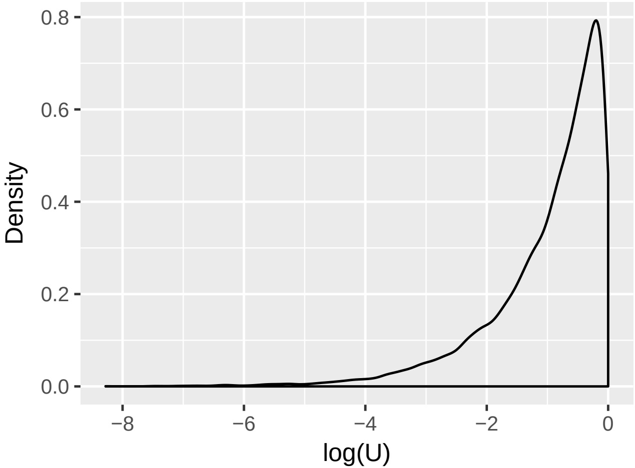
Could we construct the simplest possible distribution defined on positive values only?

Let's start with a uniform random variable.

$$U \sim \text{Uniform}(0, 1)$$

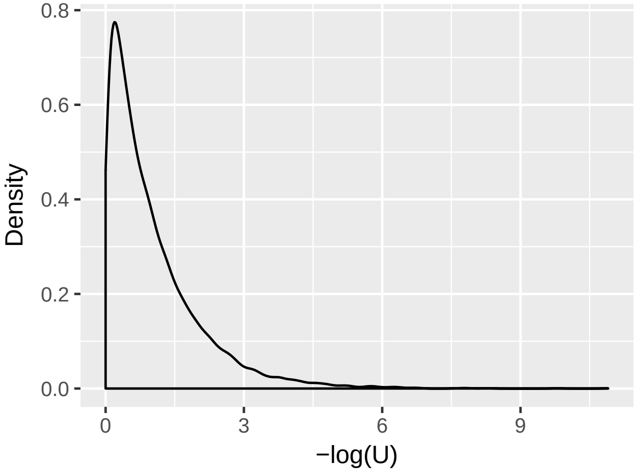


We could transform this to be defined on all negative values.



Then we could make this positive.

$$X \sim -\log(U)$$



This is the **unit exponential distribution**!

This is the **unit exponential distribution**!

$$X \sim -\log U$$

This is the **unit exponential distribution**!

$$
\begin{aligned}
X &\sim -\log U \\
-X &\sim \log U
\end{aligned}
$$

This is the **unit exponential distribution**!

$$
\begin{aligned}
X &\sim -\log U \\
-X &\sim \log U \\
e^{-X} &\sim U
\end{aligned}
$$

This is the **unit exponential distribution**!

$$
\begin{aligned}
X &\sim -\log U \\
-X &\sim \log U \\
e^{-X} &\sim U \\
1 - e^{-X} &\sim U
\end{aligned}
$$

This is the **unit exponential distribution**!

$$
\begin{aligned}
X &\sim -\log U \\
-X &\sim \log U \\
e^{-X} &\sim U \\
1 - e^{-X} &\sim U
\end{aligned}
$$

Is this starting to look like the Exponential CDF, $1 - e^{-\lambda x}$?

We can stretch it out by a **scale parameter** $\frac{1}{\lambda}$

$$X \sim -\frac{1}{\lambda}\log(U)$$

We can stretch it out by a **scale parameter** $\frac{1}{\lambda}$

$$X \sim -\frac{1}{\lambda}\log(U)$$
$$-\lambda X \sim \log(U)$$

We can stretch it out by a **scale parameter** $\frac{1}{\lambda}$

$$
\begin{aligned}
X &\sim -\frac{1}{\lambda}\log(U) \\
-\lambda X &\sim \log(U) \\
e^{-\lambda X} &\sim U
\end{aligned}
$$

We can stretch it out by a **scale parameter** $\frac{1}{\lambda}$

$$
\begin{aligned}
X &\sim -\frac{1}{\lambda}\log(U) \\
-\lambda X &\sim \log(U) \\
e^{-\lambda X} &\sim U \\
1 - e^{-\lambda X} &\sim U
\end{aligned}
$$

We can stretch it out by a **scale parameter** $\frac{1}{\lambda}$

$$\begin{aligned} X &\sim -\frac{1}{\lambda}\log(U) \\ -\lambda X &\sim \log(U) \\ e^{-\lambda X} &\sim U \\ 1 - e^{-\lambda X} &\sim U \end{aligned}$$

We have the Exponential CDF!

$$F_X(x) = 1 - e^{-\lambda x}$$

Exponential distribution

$$T \sim \text{Exponential}(\lambda)$$

PDF

$$f(t) = \lambda e^{-\lambda t}$$

CDF

$$F(t) = 1 - e^{-\lambda t}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) =$$

Exponential distribution

$$T \sim \text{Exponential}(\lambda)$$

PDF

$$f(t) = \lambda e^{-\lambda t}$$

CDF

$$F(t) = 1 - e^{-\lambda t}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) =$$

Exponential distribution

$$T \sim \text{Exponential}(\lambda)$$

PDF

$$f(t) = \lambda e^{-\lambda t}$$

CDF

$$F(t) = 1 - e^{-\lambda t}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-\lambda t}$$

Hazard function:

Exponential distribution

$$T \sim \text{Exponential}(\lambda)$$

PDF

$$f(t) = \lambda e^{-\lambda t}$$

CDF

$$F(t) = 1 - e^{-\lambda t}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-\lambda t}$$

Hazard function: Risk of event at t given survival up to t

Exponential distribution

$$T \sim \text{Exponential}(\lambda)$$

PDF

$$f(t) = \lambda e^{-\lambda t}$$

CDF

$$F(t) = 1 - e^{-\lambda t}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-\lambda t}$$

Hazard function: Risk of event at t given survival up to t

$$h(t) =$$

Exponential distribution

$$T \sim \text{Exponential}(\lambda)$$

PDF

$$f(t) = \lambda e^{-\lambda t}$$

CDF

$$F(t) = 1 - e^{-\lambda t}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-\lambda t}$$

Hazard function: Risk of event at t given survival up to t

$$h(t) = \frac{f(t)}{S(t)} =$$

Exponential distribution

$$T \sim \text{Exponential}(\lambda)$$

PDF

$$f(t) = \lambda e^{-\lambda t}$$

CDF

$$F(t) = 1 - e^{-\lambda t}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-\lambda t}$$

Hazard function: Risk of event at t given survival up to t

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{}$$

Exponential distribution

$$T \sim \text{Exponential}(\lambda)$$

PDF

$$f(t) = \lambda e^{-\lambda t}$$

CDF

$$F(t) = 1 - e^{-\lambda t}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-\lambda t}$$

Hazard function: Risk of event at t given survival up to t

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} =$$

Exponential distribution

$$T \sim \text{Exponential}(\lambda)$$

PDF

$$f(t) = \lambda e^{-\lambda t}$$

CDF

$$F(t) = 1 - e^{-\lambda t}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-\lambda t}$$

Hazard function: Risk of event at t given survival up to t

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

Exponential distribution

$$T \sim \text{Exponential}(\lambda)$$

PDF

$$f(t) = \lambda e^{-\lambda t}$$

CDF

$$F(t) = 1 - e^{-\lambda t}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-\lambda t}$$

Hazard function: Risk of event at t given survival up to t

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

Not a function of t! The hazard is **constant**.

Modeling with covariates

Suppose we want to allow the hazard to vary by some set of predictors.

Modeling with covariates

Suppose we want to allow the hazard to vary by some set of predictors.

Then, we can assume a **proportional hazards** model.

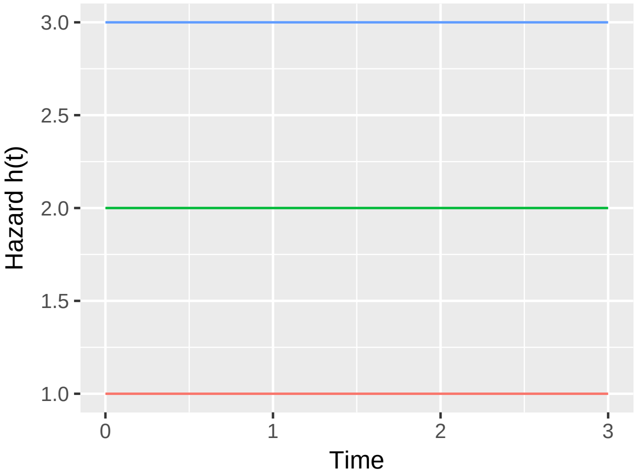
Modeling with covariates

Suppose we want to allow the hazard to vary by some set of predictors.

Then, we can assume a **proportional hazards** model.

$$h(t \mid x) = \underbrace{h_0(t)}_{\text{Baseline hazard}} \underbrace{e^{x\beta}}_{\text{Hazard ratio}}$$

Exponential hazards



Fitting an Exponential model in Zelig

Zelig is an R package designed to make everything we do in class easier.

Note the Zelig workflow overview.

We will use the Zelig-Exponential.

Zelig example: Lung cancer survival

We will walk through the example using data on lung cancer survival

```
> library(survival)
> data(lung)
> head(lung)
  inst time status age sex ph.ecog ph.karno pat.karno meal.cal wt.loss
1    3  306      2  74   1       1       90       100     1175      NA
2    3  455      2  68   1       0       90        90     1225      15
3    3 1010      1  56   1       0       90        90       NA      15
4    5  210      2  57   1       1       90        60     1150      11
5    1  883      2  60   1       0      100        90       NA       0
6   12 1022      1  74   1       1       50        80      513       0
lung <- mutate(lung, event = as.numeric(status == 2))
```

Variable definitions: Lung cancer survival

```
?lung
inst: Institution code
time: Survival time in days
status: censoring status 1=censored, 2=dead
age: Age in years
sex: Male=1 Female=2
ph.ecog: ECOG performance score (0=good 5=dead)
ph.karno: Karnofsky performance score (bad=0-good=100) rated by physician
pat.karno: Karnofsky performance score as rated by patient
meal.cal: Calories consumed at meals
wt.loss: Weight loss in last six months
```

Zelig step 1: Fit a model

```
fit <- zelig(Surv(time, event) ~ age + sex,
             model = "exp",
             data = lung)
```

Zelig step 1: Fit a model

```
> summary(fit)
Model:

Call:
z5$zelig(formula = Surv(time, event) ~ age + sex, data = lung)
              Value Std. Error     z        p
(Intercept)  6.3597    0.63547 10.01 1.41e-23
age         -0.0156    0.00911 -1.72 8.63e-02
sex          0.4809    0.16709  2.88 4.00e-03

Scale fixed at 1

Exponential distribution
Loglik(model)= -1156.1   Loglik(intercept only)= -1162.3
Chisq= 12.48 on 2 degrees of freedom, p= 0.002
Number of Newton-Raphson Iterations: 4
n= 228

Next step: Use 'setx' method
```

Zelig step 2: Use `setx` to set covariates of interest

```
men <- setx(fit, age = 50, sex = 1)
women <- setx(fit, age = 50, sex = 2)
```

Zelig step 2: Use `setx` to set covariates of interest

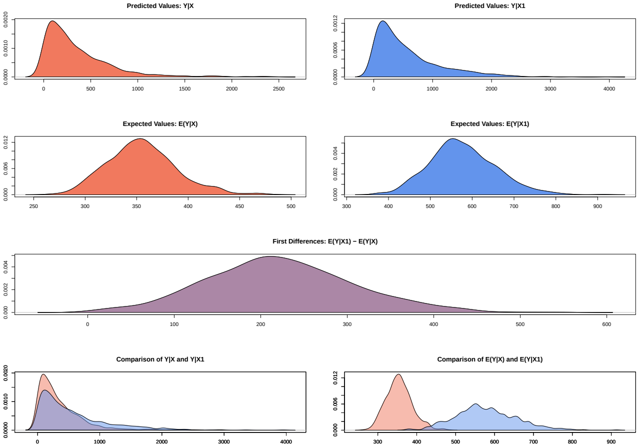
```
> men
setx:
  (Intercept) age sex
1           1  50   1

Next step: Use 'sim' method
> women
setx:
  (Intercept) age sex
1           1  50   2

Next step: Use 'sim' method
```

Zelig step 3: Use `sim` to simulate quantities of interest

```
> sims <- sim(obj = fit, x = men, x1 = women)
> summary(sims)

 sim x :
 -----
ev
     mean       sd      50%     2.5%   97.5%
1 355.086 33.63733 353.5258 296.6169 428.758
pv
        mean       sd     50%     2.5%    97.5%
[1,] 351.414 361.6174 242.511 7.082744 1357.005

 sim x1 :
 -----
ev
      mean      sd     50%     2.5%    97.5%
1 577.5684 78.5113 571.178 438.4341 743.9957
pv
         mean       sd      50%    2.5%   97.5%
[1,] 562.8317 550.6102 382.9658 11.5627 2016.61
fd
      mean      sd      50%     2.5%    97.5%
1 222.4824 85.0493 217.0278 61.08082 396.5632
```

Zelig step 4: Use graph to plot simulation results

```
pdf("ZeligFigures.pdf",
    height = 5, width = 7)
plot(sims)
dev.off()
```



Summarizing Zelig

Estimate your model:

Summarizing Zelig

Estimate your model:

```
#install.packages("Zelig")
require(Zelig)
fit <- zelig(Surv(time, event) ~ age + sex,
             model = "exp",
             data = lung)
```

Summarizing Zelig

Estimate your model:

```
#install.packages("Zelig")
require(Zelig)
fit <- zelig(Surv(time, event) ~ age + sex,
             model = "exp",
             data = lung)
```

Set your covariates:

Summarizing Zelig

Estimate your model:

```
#install.packages("Zelig")
require(Zelig)
fit <- zelig(Surv(time, event) ~ age + sex,
             model = "exp",
             data = lung)
```

Set your covariates:

```
men <- setx(fit, sex = 1, fn = mean)
women <- setx(fit, sex = 2, fn = mean)
```

Summarizing Zelig

Estimate your model:

```
#install.packages("Zelig")
require(Zelig)
fit <- zelig(Surv(time, event) ~ age + sex,
             model = "exp",
             data = lung)
```

Set your covariates:

```
men <- setx(fit, sex = 1, fn = mean)
women <- setx(fit, sex = 2, fn = mean)
```

Simulate your QOI:

Summarizing Zelig

Estimate your model:

```
#install.packages("Zelig")
require(Zelig)
fit <- zelig(Surv(time, event) ~ age + sex,
             model = "exp",
             data = lung)
```

Set your covariates:

```
men <- setx(fit, sex = 1, fn = mean)
women <- setx(fit, sex = 2, fn = mean)
```

Simulate your QOI:

```
sims <- sim(obj = fit, x = men, x1 = women)
```

Summarizing Zelig

Estimate your model:

```
#install.packages("Zelig")
require(Zelig)
fit <- zelig(Surv(time, event) ~ age + sex,
             model = "exp",
             data = lung)
```

Set your covariates:

```
men <- setx(fit, sex = 1, fn = mean)
women <- setx(fit, sex = 2, fn = mean)
```

Simulate your QOI:

```
sims <- sim(obj = fit, x = men, x1 = women)
```

Plot:

Summarizing Zelig

Estimate your model:

```
#install.packages("Zelig")
require(Zelig)
fit <- zelig(Surv(time, event) ~ age + sex,
             model = "exp",
             data = lung)
```

Set your covariates:

```
men <- setx(fit, sex = 1, fn = mean)
women <- setx(fit, sex = 2, fn = mean)
```

Simulate your QOI:

```
sims <- sim(obj = fit, x = men, x1 = women)
```

Plot:

```
plot(sims)
```

Fitting an Exponential with survreg

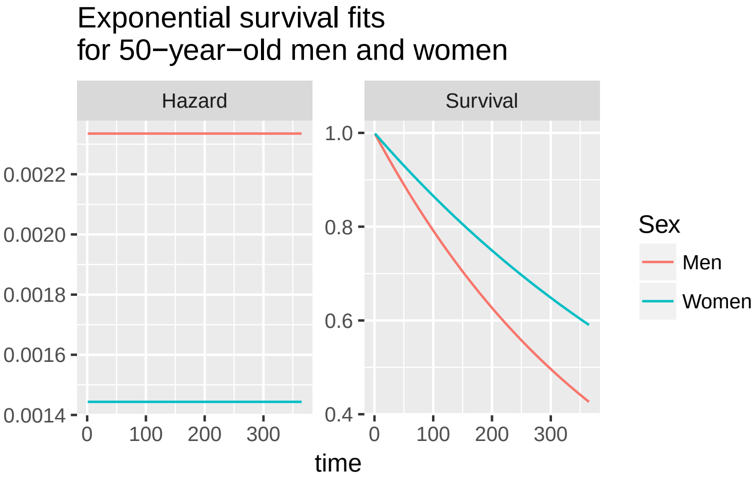
```
> library(survival)
> fit <- survreg(Surv(time, event) ~ age + sex,
+                dist = "exponential",
+                data = lung)
> summary(fit)

Call:
survreg(formula = Surv(time, event) ~ age + sex, data = lung,
    dist = "exponential")
              Value Std. Error     z        p
(Intercept)  6.3597    0.63547 10.01 1.41e-23
age         -0.0156    0.00911 -1.72 8.63e-02
sex          0.4809    0.16709  2.88 4.00e-03

Scale fixed at 1

Exponential distribution
Loglik(model)= -1156.1   Loglik(intercept only)= -1162.3
Chisq= 12.48 on 2 degrees of freedom, p= 0.002
Number of Newton-Raphson Iterations: 4
n= 228
```

Interpreting hazard ratios

$$h(t \mid x) = h_0(t)e^{-x\beta}$$

```
> exp(-coef(fit))
(Intercept)         age         sex
      0.002       1.016       0.618
```

Plotting survival curves



Plotting survival curves

How we made the previous slide:

```
data.frame(t = seq(.5,20,.5)) %>%
  mutate(Men.Hazard = lambda[1],
         Women.Hazard = lambda[2],
         Men.Survival = exp(-lambda[1]*t),
         Women.Survival = exp(-lambda[2]*t)) %>%
  melt(id = "t") %>%
  separate(variable, into = c("Sex","QOI")) %>%
  ggplot(aes(x = t, y = value, color = Sex)) +
  geom_line() +
  facet_wrap(~QOI, scales = "free") + ylab("") + xlab("time") +
  ggtitle("Exponential survival fits, for 50-year-old men and women") +
  ggsave("ExpoFit.pdf",
         height = 3, width = 5)
```

Scales and rates

The exponential is almost always parameterized with a **rate** λ.

Scales and rates

The exponential is almost always parameterized with a **rate** λ.

But, it could just as well be defined in terms of a **scale** $\theta = \frac{1}{\lambda}$

Scales and rates

The exponential is almost always parameterized with a **rate** λ.

But, it could just as well be defined in terms of a **scale** $\theta = \frac{1}{\lambda}$

Rate parameterization	**Scale parameterization**
$E(T) = \frac{1}{\lambda}$	$E(T) =$

Scales and rates

The exponential is almost always parameterized with a **rate** λ.

But, it could just as well be defined in terms of a **scale** $\theta = \frac{1}{\lambda}$

Rate parameterization	**Scale parameterization**
$E(T) = \frac{1}{\lambda}$	$E(T) = \theta$

Scales and rates

The exponential is almost always parameterized with a **rate** λ.

But, it could just as well be defined in terms of a **scale** $\theta = \frac{1}{\lambda}$

Rate parameterization	Scale parameterization
$E(T) = \frac{1}{\lambda}$	$E(T) = \theta$
$f(T) = \lambda e^{-\lambda x}$	

Scales and rates

The exponential is almost always parameterized with a **rate** λ.

But, it could just as well be defined in terms of a **scale** $\theta = \frac{1}{\lambda}$

Rate parameterization	**Scale parameterization**
$E(T) = \frac{1}{\lambda}$	$E(T) = \theta$
$f(T) = \lambda e^{-\lambda x}$	

Scales and rates

The exponential is almost always parameterized with a **rate** λ.

But, it could just as well be defined in terms of a **scale** $\theta = \frac{1}{\lambda}$

Rate parameterization	Scale parameterization
$E(T) = \frac{1}{\lambda}$	$E(T) = \theta$
$f(T) = \lambda e^{-\lambda x}$	$f(T) = \frac{1}{\theta} e^{-\frac{1}{\theta} x}$

Scales and rates

The exponential is almost always parameterized with a **rate** λ.

But, it could just as well be defined in terms of a **scale** $\theta = \frac{1}{\lambda}$

Rate parameterization	**Scale parameterization**
$E(T) = \frac{1}{\lambda}$	$E(T) = \theta$
$f(T) = \lambda e^{-\lambda x}$ As rate grows, expected waiting time shrinks	$f(T) = \frac{1}{\theta} e^{-\frac{1}{\theta} x}$

Scales and rates

The exponential is almost always parameterized with a **rate** λ.

But, it could just as well be defined in terms of a **scale** $\theta = \frac{1}{\lambda}$

Rate parameterization	**Scale parameterization**
$E(T) = \frac{1}{\lambda}$	$E(T) = \theta$
$f(T) = \lambda e^{-\lambda x}$	$f(T) = \frac{1}{\theta} e^{-\frac{1}{\theta} x}$
As rate grows, expected waiting time shrinks	As scale grows, expected waiting time

Scales and rates

The exponential is almost always parameterized with a **rate** λ.

But, it could just as well be defined in terms of a **scale** $\theta = \frac{1}{\lambda}$

Rate parameterization	Scale parameterization
$E(T) = \frac{1}{\lambda}$	$E(T) = \theta$
$f(T) = \lambda e^{-\lambda x}$	$f(T) = \frac{1}{\theta} e^{-\frac{1}{\theta} x}$
As rate grows, expected waiting time shrinks	As scale grows, expected waiting time grows

Scales and rates

The exponential is almost always parameterized with a **rate** λ.

But, it could just as well be defined in terms of a **scale** $\theta = \frac{1}{\lambda}$

Rate parameterization	**Scale parameterization**
$E(T) = \frac{1}{\lambda}$	$E(T) = \theta$
$f(T) = \lambda e^{-\lambda x}$ As rate grows, expected waiting time shrinks	$f(T) = \frac{1}{\theta} e^{-\frac{1}{\theta} x}$ As scale grows, expected waiting time grows

In general, you have to be careful with the parameterization of survival distributions.

Weibull distribution

$$T \sim \text{Weibull}(\alpha, \lambda)$$

PDF [2]

$$f(t) = t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}$$

CDF

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}}$$

Survival function

$$S(t) = P(T > t) =$$

[2] I have used the rate parameterization for λ; in lecture slides Brandon uses the scale parameterization.

Weibull distribution

$$T \sim \text{Weibull}(\alpha, \lambda)$$

PDF [2]

$$f(t) = t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}$$

CDF

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) =$$

[2] I have used the rate parameterization for λ; in lecture slides Brandon uses the scale parameterization.

Weibull distribution

$$T \sim \text{Weibull}(\alpha, \lambda)$$

PDF [2]

$$f(t) = t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}$$

CDF

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) =$$

[2] I have used the rate parameterization for λ; in lecture slides Brandon uses the scale parameterization.

Weibull distribution

$$T \sim \text{Weibull}(\alpha, \lambda)$$

PDF [2]

$$f(t) = t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}$$

CDF

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-(\lambda t)^{\alpha}}$$

Hazard function:

---

[2] I have used the rate parameterization for λ; in lecture slides Brandon uses the scale parameterization.

Weibull distribution

$$T \sim \text{Weibull}(\alpha, \lambda)$$

PDF [2]

$$f(t) = t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}$$

CDF

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-(\lambda t)^{\alpha}}$$

Hazard function: Risk of event at t given survival up to t

[2] I have used the rate parameterization for λ; in lecture slides Brandon uses the scale parameterization.

Weibull distribution

$$T \sim \text{Weibull}(\alpha, \lambda)$$

PDF [2]

$$f(t) = t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}$$

CDF

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-(\lambda t)^{\alpha}}$$

Hazard function: Risk of event at t given survival up to t

$$h(t) =$$

[2] I have used the rate parameterization for λ; in lecture slides Brandon uses the scale parameterization.

Weibull distribution

$$T \sim \text{Weibull}(\alpha, \lambda)$$

PDF [2]

$$f(t) = t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}$$

CDF

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-(\lambda t)^{\alpha}}$$

Hazard function: Risk of event at t given survival up to t

$$h(t) = \frac{f(t)}{S(t)} =$$

[2] I have used the rate parameterization for λ; in lecture slides Brandon uses the scale parameterization.

Weibull distribution

$$T \sim \text{Weibull}(\alpha, \lambda)$$

PDF [2]

$$f(t) = t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}$$

CDF

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-(\lambda t)^{\alpha}}$$

Hazard function: Risk of event at t given survival up to t

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{}$$

[2] I have used the rate parameterization for λ; in lecture slides Brandon uses the scale parameterization.

Weibull distribution

$$T \sim \text{Weibull}(\alpha, \lambda)$$

PDF [2]

$$f(t) = t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}$$

CDF

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-(\lambda t)^{\alpha}}$$

Hazard function: Risk of event at t given survival up to t

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{e^{-(\lambda t)^{\alpha}}} =$$

[2] I have used the rate parameterization for λ; in lecture slides Brandon uses the scale parameterization.

Weibull distribution

$$T \sim \text{Weibull}(\alpha, \lambda)$$

PDF [2]

$$f(t) = t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}$$

CDF

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = e^{-(\lambda t)^{\alpha}}$$

Hazard function: Risk of event at t given survival up to t

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{e^{-(\lambda t)^{\alpha}}} = t^{\alpha-1}\alpha\lambda^{\alpha}$$

[2] I have used the rate parameterization for λ; in lecture slides Brandon uses the scale parameterization.

Weibull distribution

$$h(t) =$$

Weibull distribution

$$h(t) = \frac{f(t)}{S(t)} =$$

Weibull distribution

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{}$$

Weibull distribution

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{e^{-(\lambda t)^{\alpha}}} =$$

Weibull distribution

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{e^{-(\lambda t)^{\alpha}}} = t^{\alpha-1}\alpha\lambda^{\alpha}$$

Weibull distribution

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{e^{-(\lambda t)^{\alpha}}} = t^{\alpha-1}\alpha\lambda^{\alpha}$$

The hazard increases with t when α

Weibull distribution

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{e^{-(\lambda t)^{\alpha}}} = t^{\alpha-1}\alpha\lambda^{\alpha}$$

The hazard increases with t when $\alpha > 1$

Weibull distribution

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{e^{-(\lambda t)^{\alpha}}} = t^{\alpha-1}\alpha\lambda^{\alpha}$$

The hazard increases with t when $\alpha > 1$
The hazard decreases with t when α

Weibull distribution

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{e^{-(\lambda t)^{\alpha}}} = t^{\alpha-1}\alpha\lambda^{\alpha}$$

The hazard increases with t when $\alpha > 1$
The hazard decreases with t when $\alpha < 1$

Weibull distribution

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{e^{-(\lambda t)^{\alpha}}} = t^{\alpha-1}\alpha\lambda^{\alpha}$$

The hazard increases with t when $\alpha > 1$
The hazard decreases with t when $\alpha < 1$
The hazard is constant over t when α

Weibull distribution

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{e^{-(\lambda t)^{\alpha}}} = t^{\alpha-1}\alpha\lambda^{\alpha}$$

The hazard increases with t when $\alpha > 1$
The hazard decreases with t when $\alpha < 1$
The hazard is constant over t when $\alpha = 1$

Weibull distribution

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{e^{-(\lambda t)^{\alpha}}} = t^{\alpha-1}\alpha\lambda^{\alpha}$$

The hazard increases with t when $\alpha > 1$
The hazard decreases with t when $\alpha < 1$
The hazard is constant over t when $\alpha = 1$
In that case, it's the exponential!

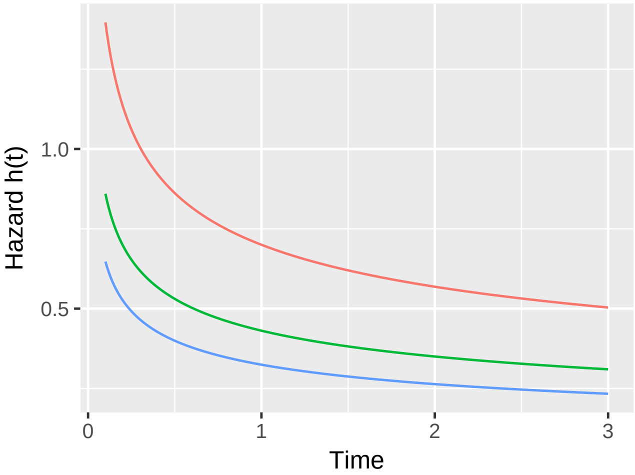
Weibull distribution

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\alpha-1}\alpha\lambda^{\alpha}e^{-(\lambda t)^{\alpha}}}{e^{-(\lambda t)^{\alpha}}} = t^{\alpha-1}\alpha\lambda^{\alpha}$$

The hazard increases with t when $\alpha > 1$
The hazard decreases with t when $\alpha < 1$
The hazard is constant over t when $\alpha = 1$
In that case, it's the exponential!

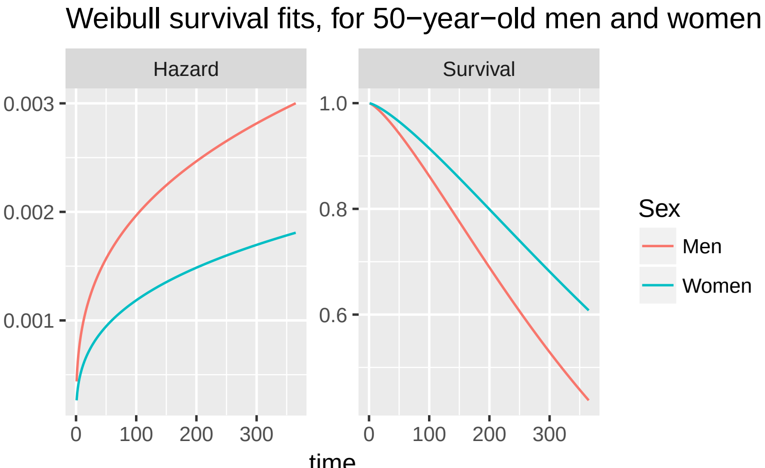
$$h(t \mid \alpha = 1) = t^{\alpha-1}\alpha\lambda^{\alpha} = t^{1-1}1\lambda^{1} = \lambda$$

Weibull hazards



Fitting a Weibull model

```
## Fitting a Weibull model
fit <- survreg(Surv(time, event) ~ age + sex,
               dist = "weibull",
               data = lung)
```



Lognormal distribution

$$T \sim \text{LogNormal}(\mu, \sigma^2) \sim e^Z \text{ (where } Z \sim N(\mu, \sigma^2))$$

$$f(t) = \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln(Y_i) - \mu)^2}{2\sigma^2}\right)$$

CDF

$$F(t) = \int_0^t f(x)dx = \text{ugly formula}$$

Survival function

$$S(t) = P(T > t) =$$

Lognormal distribution

$$T \sim \text{LogNormal}(\mu, \sigma^2) \sim e^Z \text{ (where } Z \sim N(\mu, \sigma^2))$$

$$f(t) = \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln(Y_i) - \mu)^2}{2\sigma^2}\right)$$

CDF

$$F(t) = \int_0^t f(x)dx = \text{ugly formula}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) =$$

Lognormal distribution

$$T \sim \text{LogNormal}(\mu, \sigma^2) \sim e^Z \text{ (where } Z \sim N(\mu, \sigma^2)\text{)}$$

$$f(t) = \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln(Y_i) - \mu)^2}{2\sigma^2}\right)$$

CDF

$$F(t) = \int_0^t f(x)dx = \text{ugly formula}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = \text{ugly formula}$$

Hazard function:

Lognormal distribution

$$T \sim \text{LogNormal}(\mu, \sigma^2) \sim e^Z \text{ (where } Z \sim N(\mu, \sigma^2))$$

$$f(t) = \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln(Y_i) - \mu)^2}{2\sigma^2}\right)$$

CDF

$$F(t) = \int_0^t f(x)dx = \text{ugly formula}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = \text{ugly formula}$$

Hazard function: Risk of event at t given survival up to t

Lognormal distribution

$$T \sim \text{LogNormal}(\mu, \sigma^2) \sim e^Z \text{ (where } Z \sim N(\mu, \sigma^2))$$

$$f(t) = \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln(Y_i) - \mu)^2}{2\sigma^2}\right)$$

CDF

$$F(t) = \int_0^t f(x)dx = \text{ugly formula}$$

Survival function

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F_T(t) = \text{ugly formula}$$

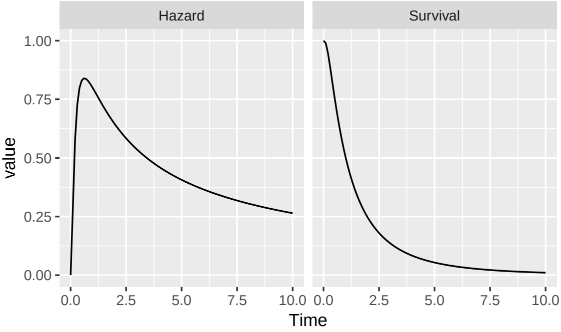
Hazard function: Risk of event at t given survival up to t

$$h(t) = \frac{f(t)}{S(t)} = \text{ugly formula}$$

Fitting a Lognormal

```
fit <- survreg(Surv(time, event) ~ age + sex,
               dist = "lognormal",
               data = lung)
```

NOTE: This figure doesn't correspond to the model above - just an example of a LogNormal



Gompertz distribution

$$f(t) = b\eta e^{bt} e^{\eta} \exp(-\eta e^{bt})$$

Gompertz distribution

$$f(t) = b\eta e^{bt} e^{\eta} \exp(-\eta e^{bt})$$
$$F(t) = 1 - \exp\left(-\eta\left(e^{bt} - 1\right)\right)$$

Gompertz distribution

$$
\begin{aligned}
f(t) &= b\eta e^{bt} e^{\eta} \exp(-\eta e^{bt}) \\
F(t) &= 1 - \exp\left(-\eta\left(e^{bt} - 1\right)\right) \\
h(t) &= \frac{f(t)}{S(t)}
\end{aligned}
$$

Gompertz distribution

$$
\begin{aligned}
f(t) &= b\eta e^{bt} e^{\eta} \exp(-\eta e^{bt}) \\
F(t) &= 1 - \exp\left(-\eta\left(e^{bt} - 1\right)\right) \\
h(t) &= \frac{f(t)}{S(t)} \\
&= \frac{b\eta e^{bt} e^{\eta} \exp(-\eta e^{bt})}{\exp\left(-\eta\left(e^{bt} - 1\right)\right)}
\end{aligned}
$$

Gompertz distribution

$$
\begin{aligned}
f(t) &= b\eta e^{bt} e^{\eta} \exp(-\eta e^{bt}) \\
F(t) &= 1 - \exp\left(-\eta\left(e^{bt} - 1\right)\right) \\
h(t) &= \frac{f(t)}{S(t)} \\
&= \frac{b\eta e^{bt} e^{\eta} \exp(-\eta e^{bt})}{\exp\left(-\eta\left(e^{bt} - 1\right)\right)} \\
&= b\eta e^{bt} e^{\eta}
\end{aligned}
$$

Gompertz distribution

$$
\begin{aligned}
f(t) &= b\eta e^{bt} e^{\eta} \exp(-\eta e^{bt}) \\
F(t) &= 1 - \exp\left(-\eta\left(e^{bt} - 1\right)\right) \\
h(t) &= \frac{f(t)}{S(t)} \\
&= \frac{b\eta e^{bt} e^{\eta} \exp(-\eta e^{bt})}{\exp\left(-\eta\left(e^{bt} - 1\right)\right)} \\
&= b\eta e^{bt} e^{\eta} \\
\log[h(t)] &= \underbrace{(\log(b) + \log(\eta) + \eta)}_{\text{Intercept}} + \underbrace{b}_{\text{Slope}} t
\end{aligned}
$$

Gompertz distribution

$$
\begin{aligned}
f(t) &= b\eta e^{bt} e^{\eta} \exp(-\eta e^{bt}) \\
F(t) &= 1 - \exp\left(-\eta\left(e^{bt} - 1\right)\right) \\
h(t) &= \frac{f(t)}{S(t)} \\
&= \frac{b\eta e^{bt} e^{\eta} \exp(-\eta e^{bt})}{\exp\left(-\eta\left(e^{bt} - 1\right)\right)} \\
&= b\eta e^{bt} e^{\eta} \\
\log[h(t)] &= \underbrace{(\log(b) + \log(\eta) + \eta)}_{\text{Intercept}} + \underbrace{b}_{\text{Slope}} t \\
&= \alpha + \beta t
\end{aligned}
$$

Gompertz distribution

$$
\begin{aligned}
f(t) &= b\eta e^{bt} e^{\eta} \exp(-\eta e^{bt}) \\
F(t) &= 1 - \exp\left(-\eta\left(e^{bt} - 1\right)\right) \\
h(t) &= \frac{f(t)}{S(t)} \\
&= \frac{b\eta e^{bt} e^{\eta} \exp(-\eta e^{bt})}{\exp\left(-\eta\left(e^{bt} - 1\right)\right)} \\
&= b\eta e^{bt} e^{\eta} \\
\log[h(t)] &= \underbrace{(\log(b) + \log(\eta) + \eta)}_{\text{Intercept}} + \underbrace{b}_{\text{Slope}} t \\
&= \alpha + \beta t
\end{aligned}
$$

The log of the hazard function is linear in time!
This is why people like the Gompertz.

Gompertz distribution

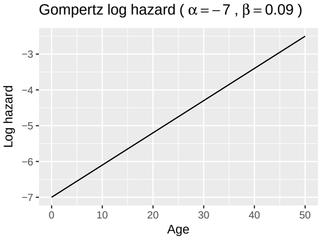
Gompertz hazard with $\alpha = -7, \beta = .09$

$$\log[h(t)] = \alpha + \beta t, \quad h(t) = \exp(\alpha + \beta t)$$

Gompertz distribution

Gompertz hazard with $\alpha = -7, \beta = .09$

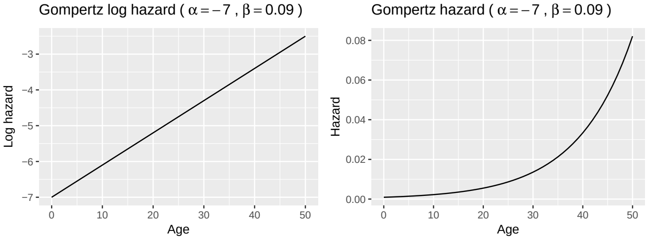
$$\log[h(t)] = \alpha + \beta t, \quad h(t) = \exp(\alpha + \beta t)$$



Gompertz distribution

Gompertz hazard with $\alpha = -7, \beta = .09$

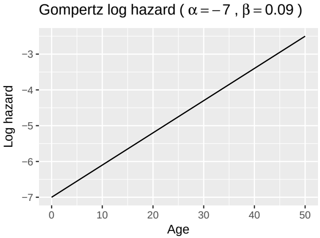
$$\log[h(t)] = \alpha + \beta t, \quad h(t) = \exp(\alpha + \beta t)$$

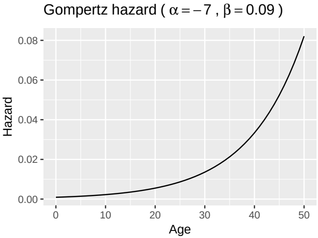


Gompertz distribution

Gompertz hazard with $\alpha = -7, \beta = .09$

$$\log[h(t)] = \alpha + \beta t, \quad h(t) = \exp(\alpha + \beta t)$$





Note: Example motivated by U.S. mortality; see German Rodriguez's example here.

As I said at the beginning, all of the survival models above have the form:

$$\underbrace{h_i(t)}_{\text{Hazard function}} = \underbrace{h_0(t)}_{\text{Baseline hazard}} \underbrace{e^{X_i\beta}}_{\text{Hazard ratio}}$$

$$S(t) = e^{-\int_0^t h(u)du}$$

Different models allow different kinds of flexibility in the baseline hazard $h_0(t)$.

As I said at the beginning, all of the survival models above have the form:

$$\underbrace{h_i(t)}_{\text{Hazard function}} = \underbrace{h_0(t)}_{\text{Baseline hazard}} \underbrace{e^{X_i\beta}}_{\text{Hazard ratio}}$$

$$S(t) = e^{-\int_0^t h(u)du}$$

Different models allow different kinds of flexibility in the baseline hazard $h_0(t)$.

Can we model hazard ratios without any assumptions about $h_0(t)$?

Cox proportional hazards model

Then we can fit a Cox proportional hazards model!

Cox proportional hazards model

Then we can fit a Cox proportional hazards model!

To save time, I won't cover this here, but it's important and in Brandon's lecture slides.

Cox proportional hazards model

Then we can fit a Cox proportional hazards model!

To save time, I won't cover this here, but it's important and in Brandon's lecture slides.

The Cox model is fit based on the order at which people die, rather than the times, so it does not assume a baseline hazard.

Cox proportional hazards model

Then we can fit a Cox proportional hazards model!

To save time, I won't cover this here, but it's important and in Brandon's lecture slides.

The Cox model is fit based on the order at which people die, rather than the times, so it does not assume a baseline hazard.

You can fit one with `coxph()`

Outline

Most probability distributions are related!

Most probability distributions are related!

In fact, people have put together charts of them all.

Most probability distributions are related!

In fact, people have put together charts of them all.

Here's one by Larry Lemis (William and Mary)

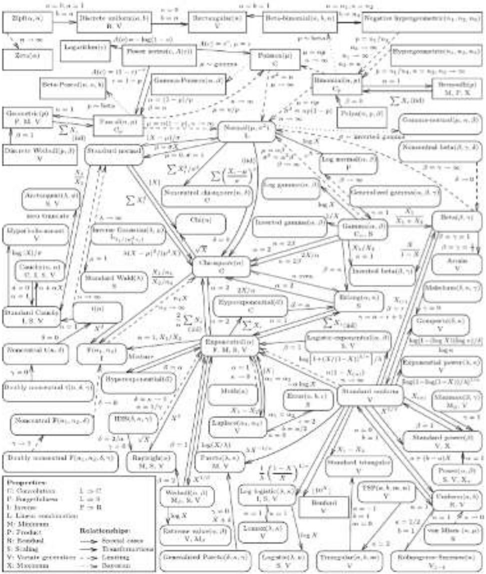


Figure 1. Univariate distribution relationships.

[link]

We won't go into all of those.

We will explore some of these relationships in the **Poisson process**.

I'd like to take you to the wilderness of the Sierra Nevada mountains, to one of my favorite places: Rae Lakes.

PC: http://wilderness.org/30-prettiest-lakes-wildlands

Imagine laying out on your pad on the granite, looking up at the sky.

We will count shooting stars and record the times we see them.[3]

[3] Thanks to William Chen for the shooting stars example. See more at `http://www.wzchen.com/probability-cheatsheet/`.

Exponential distribution: Memoryless property

$$P(T > s + t \mid T > s) =$$

Exponential distribution: Memoryless property

$$P(T > s + t \mid T > s) = \frac{P(T > s + t)}{P(T > s)} =$$

Exponential distribution: Memoryless property

$$P(T > s + t \mid T > s) = \frac{P(T > s + t)}{P(T > s)} = \frac{S(s+t)}{S(s)}$$

$$=$$

Exponential distribution: Memoryless property

$$
\begin{aligned}
P(T > s+t \mid T > s) &= \frac{P(T > s+t)}{P(T > s)} = \frac{S(s+t)}{S(s)} \\
&= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} =
\end{aligned}
$$

Exponential distribution: Memoryless property

$$\begin{aligned}
P(T > s + t \mid T > s) &= \frac{P(T > s + t)}{P(T > s)} = \frac{S(s + t)}{S(s)} \\
&= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \frac{e^{-\lambda s} e^{-\lambda t}}{e^{-\lambda s}} \\
&=
\end{aligned}$$

Exponential distribution: Memoryless property

$$
\begin{aligned}
P(T > s + t \mid T > s) &= \frac{P(T > s + t)}{P(T > s)} = \frac{S(s+t)}{S(s)} \\
&= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \frac{e^{-\lambda s} e^{-\lambda t}}{e^{-\lambda s}} \\
&= e^{-\lambda t} =
\end{aligned}
$$

Exponential distribution: Memoryless property

$$
\begin{aligned}
P(T > s + t \mid T > s) &= \frac{P(T > s + t)}{P(T > s)} = \frac{S(s + t)}{S(s)} \\
&= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \frac{e^{-\lambda s} e^{-\lambda t}}{e^{-\lambda s}} \\
&= e^{-\lambda t} = P(T > t)
\end{aligned}
$$

Exponential distribution: Memoryless property

$$
\begin{aligned}
P(T > s + t \mid T > s) &= \frac{P(T > s + t)}{P(T > s)} = \frac{S(s + t)}{S(s)} \\
&= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \frac{e^{-\lambda s} e^{-\lambda t}}{e^{-\lambda s}} \\
&= e^{-\lambda t} = P(T > t)
\end{aligned}
$$

So, the probability of surviving an additional t years is independent of whether you have already survived s years!

Exponential ratio

Suppose

$$X_1, X_2 \overset{\text{iid}}{\sim} \text{Exponential}(1)$$

What is the distribution of:

$$\frac{X_1}{X_1 + X_2} \sim$$

Exponential ratio

Suppose

$$X_1, X_2 \overset{\text{iid}}{\sim} \text{Exponential}(1)$$

What is the distribution of:

$$\frac{X_1}{X_1 + X_2} \sim \text{Uniform}(0, X_1 + X_2)$$

Exponential ratio

Suppose

$$X_1, X_2 \stackrel{\text{iid}}{\sim} \text{Exponential}(1)$$

What is the distribution of:

$$\frac{X_1}{X_1 + X_2} \sim \text{Uniform}(0, X_1 + X_2)$$

What if X_1 and X_2 are distributed Exponential(2)?

Exponential ratio

Suppose

$$X_1, X_2 \overset{\text{iid}}{\sim} \text{Exponential}(1)$$

What is the distribution of:

$$\frac{X_1}{X_1 + X_2} \sim \text{Uniform}(0, X_1 + X_2)$$

What if X_1 and X_2 are distributed Exponential(2)? The result still holds!

Sum of exponentials

The exponential is often described as the length of time you wait until a bus comes.

Sum of exponentials

The exponential is often described as the length of time you wait until a bus comes.

What if we wanted a distribution for the time until the kth bus comes?

$$X_1, \ldots, X_k \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$$

Sum of exponentials

The exponential is often described as the length of time you wait until a bus comes.

What if we wanted a distribution for the time until the kth bus comes?

$$\begin{aligned} X_1, \ldots, X_k &\overset{\text{iid}}{\sim} \text{Exponential}(\lambda) \\ G_k &\sim X_1 + \cdots + X_k \end{aligned}$$

Sum of exponentials

The exponential is often described as the length of time you wait until a bus comes.

What if we wanted a distribution for the time until the kth bus comes?

$$X_1, \ldots, X_k \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$$
$$G_k \sim X_1 + \cdots + X_k$$

Then we say

$$G_k \sim \text{Gamma}(k, \lambda)$$

The **Gamma distribution** characterizes the wait time until the kth bus arrives.

Gamma distribution

$$G_k \sim \text{Gamma}(k, \lambda) \sim X_1 + \cdots + X_k$$

$$E(T) =$$

Gamma distribution

$$G_k \sim \text{Gamma}(k, \lambda) \sim X_1 + \cdots + X_k$$

$$E(T) = E(X_1 + \cdots + X_k) =$$

Gamma distribution

$$G_k \sim \text{Gamma}(k, \lambda) \sim X_1 + \cdots + X_k$$

$$E(T) = E(X_1 + \cdots + X_k) = E(X_1) + \cdots + E(X_k) =$$

Gamma distribution

$$G_k \sim \text{Gamma}(k, \lambda) \sim X_1 + \cdots + X_k$$

$$E(T) = E(X_1 + \cdots + X_k) = E(X_1) + \cdots + E(X_k) = \frac{k}{\lambda}$$

Gamma distribution

$$G_k \sim \text{Gamma}(k, \lambda) \sim X_1 + \cdots + X_k$$

$$E(T) = E(X_1 + \cdots + X_k) = E(X_1) + \cdots + E(X_k) = \frac{k}{\lambda}$$

$$V(T) =$$

Gamma distribution

$$G_k \sim \text{Gamma}(k, \lambda) \sim X_1 + \cdots + X_k$$

$$E(T) = E(X_1 + \cdots + X_k) = E(X_1) + \cdots + E(X_k) = \frac{k}{\lambda}$$

$$V(T) = V(X_1 + \cdots + X_k) =$$

Gamma distribution

$$G_k \sim \text{Gamma}(k, \lambda) \sim X_1 + \cdots + X_k$$

$$E(T) = E(X_1 + \cdots + X_k) = E(X_1) + \cdots + E(X_k) = \frac{k}{\lambda}$$

$$V(T) = V(X_1 + \cdots + X_k) = V(X_1) + \cdots + V(X_k) =$$

Gamma distribution

$$G_k \sim \text{Gamma}(k, \lambda) \sim X_1 + \cdots + X_k$$

$$E(T) = E(X_1 + \cdots + X_k) = E(X_1) + \cdots + E(X_k) = \frac{k}{\lambda}$$

$$V(T) = V(X_1 + \cdots + X_k) = V(X_1) + \cdots + V(X_k) = \frac{k}{\lambda^2}$$

Gamma distribution

$$G_k \sim \mathrm{Gamma}(k, \lambda) \sim X_1 + \cdots + X_k$$

$$E(T) = E(X_1 + \cdots + X_k) = E(X_1) + \cdots + E(X_k) = \frac{k}{\lambda}$$

$$V(T) = V(X_1 + \cdots + X_k) = V(X_1) + \cdots + V(X_k) = \frac{k}{\lambda^2}$$

$$f(t) = \frac{1}{\Gamma(k)} \frac{(\lambda t)^k e^{-\lambda t}}{t}$$

Gamma distribution

$$G_k \sim \text{Gamma}(k, \lambda) \sim X_1 + \cdots + X_k$$

$$E(T) = E(X_1 + \cdots + X_k) = E(X_1) + \cdots + E(X_k) = \frac{k}{\lambda}$$

$$V(T) = V(X_1 + \cdots + X_k) = V(X_1) + \cdots + V(X_k) = \frac{k}{\lambda^2}$$

$$f(t) = \frac{1}{\Gamma(k)} \frac{(\lambda t)^k e^{-\lambda t}}{t}$$

CDF is hard to write.

Poisson: Events in an interval

Suppose

$$X_1, X_2, \cdots \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$$

Then the number of events occurring in a window of length 1 follows a Poisson distribution with rate λ.

$$N \sim Pois(\lambda)$$

Poisson: Events in an interval

Suppose

$$X_1, X_2, \cdots \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$$

Then the number of events occurring in a window of length 1 follows a Poisson distribution with rate λ.

$$N \sim Pois(\lambda)$$

$$E(N) = \lambda$$

Poisson: Events in an interval

Suppose

$$X_1, X_2, \cdots \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$$

Then the number of events occurring in a window of length 1 follows a Poisson distribution with rate λ.

$$N \sim Pois(\lambda)$$

$$E(N) = \lambda$$

$$V(N) = \lambda$$

Poisson: Events in disjoint intervals

Are the number of events in disjoint intervals (i.e. $\{t \in (0,1), t \in (2,3)\}$) related?

Poisson: Events in disjoint intervals

Are the number of events in disjoint intervals (i.e. $\{t \in (0,1), t \in (2,3)\}$) related?

No! By the memoryless property of the Exponential, the wait times for events in these two periods are **independent** given the rate parameter λ.

Poisson: Intervals of non-unit length

What would you expect for the distribution of events occuring in a window of length 2?

Poisson: Intervals of non-unit length

What would you expect for the distribution of events occuring in a window of length 2?

$$E(N_T \mid T = 2) = 2\lambda, \qquad V(N_T \mid T = 2) = 2\lambda$$

In general it will be the case that the number of events in a time period of length t follows a Poisson distribution with rate λt.

$$N_t \sim \text{Poisson}(\lambda t)$$

Ratio of Gammas

$$G_a \sim \mathrm{Gamma}(a, \lambda)$$

Ratio of Gammas

$$G_a \sim \text{Gamma}(a, \lambda)$$
$$G_b \sim \text{Gamma}(b, \lambda)$$

Ratio of Gammas

$$
\begin{aligned}
G_a &\sim \text{Gamma}(a, \lambda) \\
G_b &\sim \text{Gamma}(b, \lambda) \\
B \equiv \frac{G_a}{G_a + G_b} &\sim \text{Beta}(a, b)
\end{aligned}
$$

Ratio of Gammas

$$
\begin{aligned}
G_a &\sim \text{Gamma}(a, \lambda) \\
G_b &\sim \text{Gamma}(b, \lambda) \\
B \equiv \frac{G_a}{G_a + G_b} &\sim \text{Beta}(a, b)
\end{aligned}
$$

The ratio of Gammas is a **Beta distribution**!

Uniform order statistics

What is the distribution of the time until the 5th shooting star?

Uniform order statistics

What is the distribution of the time until the 5th shooting star?

$$\text{Gamma}(5, \lambda)$$

Uniform order statistics

What is the distribution of the time until the 5th shooting star?

$$\text{Gamma}(5, \lambda)$$

What is the distribution of the time until the 20th shooting star after that?

Uniform order statistics

What is the distribution of the time until the 5th shooting star?

$$\text{Gamma}(5, \lambda)$$

What is the distribution of the time until the 20th shooting star after that?

$$\text{Gamma}(20, \lambda)$$

Uniform order statistics

What is the distribution of the time until the 5th shooting star?

$$\text{Gamma}(5, \lambda)$$

What is the distribution of the time until the 20th shooting star after that?

$$\text{Gamma}(20, \lambda)$$

These wait times are independent. What is the distribution of the proportion of time spent waiting for the 5th star?

Uniform order statistics

What is the distribution of the time until the 5th shooting star?

$$\text{Gamma}(5, \lambda)$$

What is the distribution of the time until the 20th shooting star after that?

$$\text{Gamma}(20, \lambda)$$

These wait times are independent. What is the distribution of the proportion of time spent waiting for the 5th star?

$$U_{(5)} \sim \frac{\text{Gamma}(5, \lambda)}{\text{Gamma}(5, \lambda) + \text{Gamma}(20, \lambda)} \sim \text{Beta}(5, 20)$$

Uniform order statistics

What is the distribution of the time until the 5th shooting star?

$$\text{Gamma}(5, \lambda)$$

What is the distribution of the time until the 20th shooting star after that?

$$\text{Gamma}(20, \lambda)$$

These wait times are independent. What is the distribution of the proportion of time spent waiting for the 5th star?

$$U_{(5)} \sim \frac{\text{Gamma}(5, \lambda)}{\text{Gamma}(5, \lambda) + \text{Gamma}(20, \lambda)} \sim \text{Beta}(5, 20)$$

Why do we care?

Suppose someone says to you, "I ran 100 hypothesis tests. What's the probability that the 7th-smallest p-value is less than 0.05 if nothing is really happening?"

Why do we care?

Suppose someone says to you, "I ran 100 hypothesis tests. What's the probability that the 7th-smallest p-value is less than 0.05 if nothing is really happening?"

You say...let me take you to the wilderness. We will count shooting stars.

Why do we care?

$$X_1, \ldots, X_n \sim \text{Exponential}$$

Why do we care?

$$X_1, \ldots, X_n \sim \text{Exponential}$$

$$\left\{ \frac{X_1}{\sum X_i}, \ldots, \frac{X_n}{\sum X_i} \right\} \sim \left\{ U_{(1)}, \ldots, U_{(n)} \right\}$$

where $\left\{ U_{(1)}, \ldots, U_{(n)} \right\}$ are **order statistics** that give an ordered version of $U_1, \ldots, U_n \sim \text{Uniform}(0, 1)$

Why do we care?

$$X_1, \ldots, X_n \sim \text{Exponential}$$

$$\left\{ \frac{X_1}{\sum X_i}, \ldots, \frac{X_n}{\sum X_i} \right\} \sim \left\{ U_{(1)}, \ldots, U_{(n)} \right\}$$

where $\left\{ U_{(1)}, \ldots, U_{(n)} \right\}$ are **order statistics** that give an ordered version of $U_1, \ldots, U_n \sim \text{Uniform}(0, 1)$

$$G_7 \equiv X_1 + \cdots + X_7 \sim \text{Gamma}(7, 1)$$

Why do we care?

$$X_1, \ldots, X_n \sim \text{Exponential}$$

$$\left\{ \frac{X_1}{\sum X_i}, \ldots, \frac{X_n}{\sum X_i} \right\} \sim \left\{ U_{(1)}, \ldots, U_{(n)} \right\}$$

where $\left\{ U_{(1)}, \ldots, U_{(n)} \right\}$ are **order statistics** that give an ordered version of $U_1, \ldots, U_n \sim \text{Uniform}(0, 1)$

$$G_7 \equiv X_1 + \cdots + X_7 \sim \text{Gamma}(7, 1)$$

$$G_{93} \equiv X_8 + \cdots + X_{100} \sim \text{Gamma}(93, 1)$$

Why do we care?

$$X_1, \ldots, X_n \sim \text{Exponential}$$

$$\left\{ \frac{X_1}{\sum X_i}, \ldots, \frac{X_n}{\sum X_i} \right\} \sim \left\{ U_{(1)}, \ldots, U_{(n)} \right\}$$

where $\left\{ U_{(1)}, \ldots, U_{(n)} \right\}$ are **order statistics** that give an ordered version of $U_1, \ldots, U_n \sim \text{Uniform}(0, 1)$

$$G_7 \equiv X_1 + \cdots + X_7 \sim \text{Gamma}(7, 1)$$

$$G_{93} \equiv X_8 + \cdots + X_{100} \sim \text{Gamma}(93, 1)$$

$$U_{(7)} \sim \frac{G_7}{G_7 + G_{93}} \sim \text{Beta}(7, 93)$$

Why do we care?

$$X_1, \dots, X_n \sim \text{Exponential}$$

$$\left\{ \frac{X_1}{\sum X_i}, \dots, \frac{X_n}{\sum X_i} \right\} \sim \left\{ U_{(1)}, \dots, U_{(n)} \right\}$$

where $\left\{ U_{(1)}, \dots, U_{(n)} \right\}$ are **order statistics** that give an ordered version of $U_1, \dots, U_n \sim \text{Uniform}(0,1)$

$$G_7 \equiv X_1 + \cdots + X_7 \sim \text{Gamma}(7,1)$$

$$G_{93} \equiv X_8 + \cdots + X_{100} \sim \text{Gamma}(93,1)$$

$$U_{(7)} \sim \frac{G_7}{G_7 + G_{93}} \sim \text{Beta}(7,93)$$

$$P(U_{(7)} < .05) = F_{\text{Beta}(7,93)}(.05) = 0.23$$

Why do we care?

$$X_1, \ldots, X_n \sim \text{Exponential}$$

$$\left\{ \frac{X_1}{\sum X_i}, \ldots, \frac{X_n}{\sum X_i} \right\} \sim \left\{ U_{(1)}, \ldots, U_{(n)} \right\}$$

where $\left\{ U_{(1)}, \ldots, U_{(n)} \right\}$ are **order statistics** that give an ordered version of $U_1, \ldots, U_n \sim \text{Uniform}(0, 1)$

$$G_7 \equiv X_1 + \cdots + X_7 \sim \text{Gamma}(7, 1)$$

$$G_{93} \equiv X_8 + \cdots + X_{100} \sim \text{Gamma}(93, 1)$$

$$U_{(7)} \sim \frac{G_7}{G_7 + G_{93}} \sim \text{Beta}(7, 93)$$

$$P(U_{(7)} < .05) = F_{\text{Beta}(7,93)}(.05) = 0.23$$

So, it's not that strange to see 7 p-values less than 0.05.

Why do we care?

$$X_1, \ldots, X_n \sim \text{Exponential}$$

$$\left\{ \frac{X_1}{\sum X_i}, \ldots, \frac{X_n}{\sum X_i} \right\} \sim \left\{ U_{(1)}, \ldots, U_{(n)} \right\}$$

where $\left\{ U_{(1)}, \ldots, U_{(n)} \right\}$ are **order statistics** that give an ordered version of $U_1, \ldots, U_n \sim \text{Uniform}(0, 1)$

$$G_7 \equiv X_1 + \cdots + X_7 \sim \text{Gamma}(7, 1)$$

$$G_{93} \equiv X_8 + \cdots + X_{100} \sim \text{Gamma}(93, 1)$$

$$U_{(7)} \sim \frac{G_7}{G_7 + G_{93}} \sim \text{Beta}(7, 93)$$

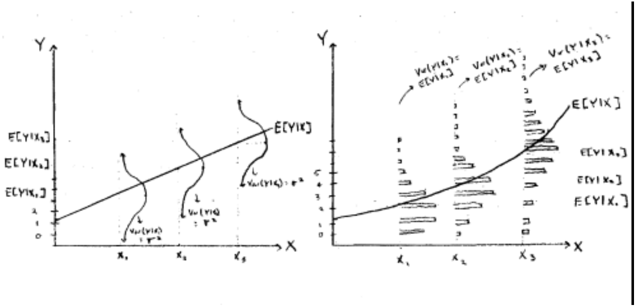
$$P(U_{(7)} < .05) = F_{\text{Beta}(7,93)}(.05) = 0.23$$

So, it's not that strange to see 7 p-values less than 0.05. And we learned this all from shooting stars!

Outline

In the Poisson distribution,

$$E(Y) = V(Y) = \lambda$$



This is fairly restrictive. Can we allow the variance to differ from the mean?

Negative Binomial: Three constructions

We can add flexibility to the Poisson model with the **Negative Binomial**.

Negative Binomial: Three constructions

We can add flexibility to the Poisson model with the **Negative Binomial**.

We will walk through three constructions of the Negative Binomial as

- The number of tails until the kth heads
- A Gamma-Poisson mixture
- A Poisson with an overdispersion parameter

Negative Binomial: Three constructions

We can add flexibility to the Poisson model with the **Negative Binomial**.

We will walk through three constructions of the Negative Binomial as

- The number of tails until the kth heads
- A Gamma-Poisson mixture
- A Poisson with an overdispersion parameter

Geometric distribution: Tails until first heads

Suppose you flip a coin until the first heads.
The number of tails until the first heads follows a **geometric** distribution.

Geometric distribution: Tails until first heads

Suppose you flip a coin until the first heads.
The number of tails until the first heads follows a **geometric** distribution.

$$Y \sim \text{Geometric}(p)$$

$$P(Y = y) = P(Y \text{ failures})P(\text{Final success}) =$$

Geometric distribution: Tails until first heads

Suppose you flip a coin until the first heads.
The number of tails until the first heads follows a **geometric** distribution.

$$Y \sim \text{Geometric}(p)$$

$$P(Y = y) = P(Y \text{ failures})P(\text{Final success}) = (1 - p)^y p$$

Negative Binomial distribution: Tails until the kth heads

$$Y_i \sim \text{NegBin}(p_i, k)$$
$$P(Y_i = y_i) = \binom{k+y-1}{k}(1-p_i)^y p_i^k$$

Note: This is a model for counts involving two parameters, so has flexibility beyond the Poisson.

Negative Binomial: Three constructions

We can add flexibility to the Poisson model with the **Negative Binomial**.

We will walk through three constructions of the Negative Binomial as

- The number of tails until the kth heads
- A Gamma-Poisson mixture
- A Poisson with an overdispersion parameter

Negative Binomial: A Gamma-Poisson mixture[4]

$$
\begin{aligned}
Y_i|\varsigma_i &\sim Poisson(\varsigma_i \lambda_i) \\
\varsigma_i &\sim \frac{1}{\theta} Gamma(\theta, 1)
\end{aligned}
$$

[4] Material adapted from lecture slides

Negative Binomial: A Gamma-Poisson mixture[4]

$$
\begin{aligned}
Y_i|\varsigma_i &\sim Poisson(\varsigma_i \lambda_i) \\
\varsigma_i &\sim \frac{1}{\theta} Gamma(\theta, 1)
\end{aligned}
$$

Note that Gamma$(\theta, 1)$ has mean θ.

[4] Material adapted from lecture slides

Negative Binomial: A Gamma-Poisson mixture[4]

$$\begin{aligned} Y_i|\varsigma_i &\sim Poisson(\varsigma_i\lambda_i) \\ \varsigma_i &\sim \frac{1}{\theta}Gamma(\theta, 1) \end{aligned}$$

Note that Gamma$(\theta, 1)$ has mean θ. This means that $\frac{1}{\theta}$Gamma(θ) has mean 1, and so Poisson$(\varsigma_i\lambda_i)$ has mean λ_i.

[4] Material adapted from lecture slides

Negative Binomial: A Gamma-Poisson mixture[5]

Using a similar approach to that described in UPM pgs. 51-52 we can derive the marginal distribution of Y as

$$Y_i \sim Negbin(\lambda_i, \theta)$$

[5] Material adapted from lecture slides

Negative Binomial: A Gamma-Poisson mixture[5]

Using a similar approach to that described in UPM pgs. 51-52 we can derive the marginal distribution of Y as

$$Y_i \sim Negbin(\lambda_i, \theta)$$

where

$$f_{nb}(y_i|\lambda_i, \theta) = \frac{\Gamma(\theta + y_i)}{y_i!\Gamma(\theta)} \frac{\lambda_i^{y_i} \theta^{\theta}}{(\lambda_i + \theta)^{\theta + y_i}}$$

[5] Material adapted from lecture slides

Negative Binomial: A Gamma-Poisson mixture[5]

Using a similar approach to that described in UPM pgs. 51-52 we can derive the marginal distribution of Y as

$$Y_i \sim Negbin(\lambda_i, \theta)$$

where

$$f_{nb}(y_i|\lambda_i, \theta) = \frac{\Gamma(\theta + y_i)}{y_i!\Gamma(\theta)} \frac{\lambda_i^{y_i} \theta^\theta}{(\lambda_i + \theta)^{\theta + y_i}}$$

Notes:

1. $\mathrm{E}[Y_i] = \lambda_i$ and $\mathrm{Var}(Y_i) = \lambda_i + \frac{\lambda_i^2}{\theta}$. What values of θ would be evidence *against* overdispersion?

[5]Material adapted from lecture slides

Negative Binomial: A Gamma-Poisson mixture[5]

Using a similar approach to that described in UPM pgs. 51-52 we can derive the marginal distribution of Y as

$$Y_i \sim Negbin(\lambda_i, \theta)$$

where

$$f_{nb}(y_i|\lambda_i, \theta) = \frac{\Gamma(\theta + y_i)}{y_i!\Gamma(\theta)} \frac{\lambda_i^{y_i}\theta^\theta}{(\lambda_i + \theta)^{\theta + y_i}}$$

Notes:

1. $\mathrm{E}[Y_i] = \lambda_i$ and $\mathrm{Var}(Y_i) = \lambda_i + \frac{\lambda_i^2}{\theta}$. What values of θ would be evidence *against* overdispersion?
2. we still have the same old systematic component: $\lambda_i = \exp(X_i\beta)$.

[5] Material adapted from lecture slides

Negative Binomial: Three constructions

We can add flexibility to the Poisson model with the **Negative Binomial**.

We will walk through three constructions of the Negative Binomial as

- The number of tails until the kth heads
- A Gamma-Poisson mixture
- A Poisson with an overdispersion parameter

Negative Binomial: Poisson with an overdispersion parameter

$$E(Y_i) = \lambda_i$$

$$V(Y_i) = \theta\lambda_i$$

$$p(y_i) = \frac{\Gamma\left(Y_i + \frac{\lambda_i}{1-\theta}\right)}{Y_i!\Gamma\left(\frac{\lambda_i}{1-\theta}\right)} \left(\frac{\lambda_i}{\lambda_i + \frac{\lambda_i}{1-\theta}}\right)^{Y_i} \left(\frac{\frac{\lambda_i}{1-\theta}}{\lambda_i + \frac{\lambda_i}{1-\theta}}\right)^{\frac{\lambda_i}{1-\theta}}$$

Harmonizing the three constructions

Failures before kth success

$$p(y) = \underbrace{\binom{k+y-1}{k}}_{\text{Part 1}} \underbrace{(1-p)^y}_{\text{Part 2}} \underbrace{p^k}_{\text{Part 3}}$$

Harmonizing the three constructions

Failures before kth success

$$p(y) = \underbrace{\binom{k+y-1}{k}}_{\text{Part 1}} \underbrace{(1-p)^y}_{\text{Part 2}} \underbrace{p^k}_{\text{Part 3}}$$

Gamma-Poisson mixture

$$p(y) = \frac{\Gamma(\theta + y_i)}{y_i!\Gamma(\theta)} \frac{\lambda_i^{y_i}\theta^\theta}{(\lambda_i + \theta)^{\theta + y_i}}$$

Harmonizing the three constructions

Failures before kth success

$$p(y) = \underbrace{\binom{k+y-1}{k}}_{\text{Part 1}} \underbrace{(1-p)^y}_{\text{Part 2}} \underbrace{p^k}_{\text{Part 3}}$$

Gamma-Poisson mixture

$$\begin{aligned} p(y) &= \frac{\Gamma(\theta + y_i)}{y_i!\Gamma(\theta)} \frac{\lambda_i^{y_i}\theta^\theta}{(\lambda_i + \theta)^{\theta + y_i}} \\ &= \underbrace{\frac{\Gamma(\theta + y_i)}{y_i!\Gamma(\theta)}}_{\text{Part 1}} \underbrace{\left(\frac{\lambda_i}{\lambda_i + \theta}\right)^{y_i}}_{\text{Part 2}} \underbrace{\left(\frac{\theta}{\lambda_i + \theta}\right)^{\theta}}_{\text{Part 3}} \end{aligned}$$

Harmonizing the three constructions

Failures before kth success

$$p(y) = \underbrace{\binom{k+y-1}{k}}_{\text{Part 1}} \underbrace{(1-p)^y}_{\text{Part 2}} \underbrace{p^k}_{\text{Part 3}}$$

Gamma-Poisson mixture

$$\begin{aligned} p(y) &= \frac{\Gamma(\theta + y_i)}{y_i!\Gamma(\theta)} \frac{\lambda_i^{y_i}\theta^\theta}{(\lambda_i + \theta)^{\theta + y_i}} \\ &= \underbrace{\frac{\Gamma(\theta + y_i)}{y_i!\Gamma(\theta)}}_{\text{Part 1}} \underbrace{\left(\frac{\lambda_i}{\lambda_i + \theta}\right)^{y_i}}_{\text{Part 2}} \underbrace{\left(\frac{\theta}{\lambda_i + \theta}\right)^{\theta}}_{\text{Part 3}} \end{aligned}$$

These are just different parameterizations of the same thing!

Harmonizing the three constructions

Failures before kth success

$$p(y) = \underbrace{\binom{k+y-1}{k}}_{\text{Part 1}} \underbrace{(1-p)^y}_{\text{Part 2}} \underbrace{p^k}_{\text{Part 3}}$$

Gamma-Poisson mixture

$$\begin{aligned} p(y) &= \frac{\Gamma(\theta + y_i)}{y_i!\Gamma(\theta)} \frac{\lambda_i^{y_i}\theta^\theta}{(\lambda_i+\theta)^{\theta+y_i}} \\ &= \underbrace{\frac{\Gamma(\theta + y_i)}{y_i!\Gamma(\theta)}}_{\text{Part 1}} \underbrace{\left(\frac{\lambda_i}{\lambda_i+\theta}\right)^{y_i}}_{\text{Part 2}} \underbrace{\left(\frac{\theta}{\lambda_i+\theta}\right)^{\theta}}_{\text{Part 3}} \end{aligned}$$

These are just different parameterizations of the same thing!

$$p = \frac{\theta}{\lambda_i + \theta}$$

Harmonizing the three constructions

Failures before kth success

$$p(y) = \underbrace{\binom{k+y-1}{k}}_{\text{Part 1}} \underbrace{(1-p)^y}_{\text{Part 2}} \underbrace{p^k}_{\text{Part 3}}$$

Gamma-Poisson mixture

$$\begin{aligned} p(y) &= \frac{\Gamma(\theta + y_i)}{y_i!\Gamma(\theta)} \frac{\lambda_i^{y_i}\theta^\theta}{(\lambda_i + \theta)^{\theta + y_i}} \\ &= \underbrace{\frac{\Gamma(\theta + y_i)}{y_i!\Gamma(\theta)}}_{\text{Part 1}} \underbrace{\left(\frac{\lambda_i}{\lambda_i + \theta}\right)^{y_i}}_{\text{Part 2}} \underbrace{\left(\frac{\theta}{\lambda_i + \theta}\right)^{\theta}}_{\text{Part 3}} \end{aligned}$$

These are just different parameterizations of the same thing!

$$p = \frac{\theta}{\lambda_i + \theta}$$

$$k = \theta$$

Harmonizing the three constructions

Failures before kth success

$$p(y) = \underbrace{\binom{k+y-1}{k}}_{\text{Part 1}} \underbrace{(1-p)^y}_{\text{Part 2}} \underbrace{p^k}_{\text{Part 3}}$$

Harmonizing the three constructions

Failures before kth success

$$p(y) = \underbrace{\binom{k+y-1}{k}}_{\text{Part 1}} \underbrace{(1-p)^y}_{\text{Part 2}} \underbrace{p^k}_{\text{Part 3}}$$

Poisson with an overdispersion parameter

$$p(y_i) = \underbrace{\frac{\Gamma(Y_i + \frac{\lambda_i}{1-\theta})}{Y_i!\Gamma(\frac{\lambda_i}{1-\theta})}}_{\text{Part 1}} \underbrace{\left(\frac{\lambda_i}{\lambda_i + \frac{\lambda_i}{1-\theta}}\right)^{Y_i}}_{\text{Part 2}} \underbrace{\left(\frac{\frac{\lambda_i}{1-\theta}}{\lambda_i + \frac{\lambda_i}{1-\theta}}\right)^{\frac{\lambda_i}{1-\theta}}}_{\text{Part 3}}$$

These are just different parameterizations of the same thing!

$$p = \frac{\frac{\lambda_i}{1-\theta}}{\lambda_i + \frac{\lambda_i}{1-\theta}}$$

Harmonizing the three constructions

Failures before kth success

$$p(y) = \underbrace{\binom{k+y-1}{k}}_{\text{Part 1}} \underbrace{(1-p)^y}_{\text{Part 2}} \underbrace{p^k}_{\text{Part 3}}$$

Poisson with an overdispersion parameter

$$p(y_i) = \underbrace{\frac{\Gamma(Y_i + \frac{\lambda_i}{1-\theta})}{Y_i!\Gamma(\frac{\lambda_i}{1-\theta})}}_{\text{Part 1}} \underbrace{\left(\frac{\lambda_i}{\lambda_i + \frac{\lambda_i}{1-\theta}}\right)^{Y_i}}_{\text{Part 2}} \underbrace{\left(\frac{\frac{\lambda_i}{1-\theta}}{\lambda_i + \frac{\lambda_i}{1-\theta}}\right)^{\frac{\lambda_i}{1-\theta}}}_{\text{Part 3}}$$

These are just different parameterizations of the same thing!

$$p = \frac{\frac{\lambda_i}{1-\theta}}{\lambda_i + \frac{\lambda_i}{1-\theta}}$$

$$k = \frac{\lambda_i}{1-\theta}$$

Outline

Zero-inflation

What if count data has a disproportionate number of 0s?

Zero-inflation

What if count data has a disproportionate number of 0s?

Brandon gave the example of someone who fishes.

Zero-inflation

What if count data has a disproportionate number of 0s?

Brandon gave the example of someone who fishes.

There's some probability of not fishing at all (catching 0 fish)

Zero-inflation

What if count data has a disproportionate number of 0s?

Brandon gave the example of someone who fishes.

There's some probability of not fishing at all (catching 0 fish)

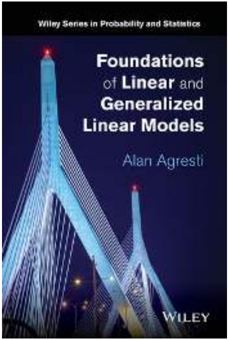
Given that you fish, there's some count distribution for the number of fish caught.

Example

Keeping with the nautical theme, we will use an example with Horseshoe crabs.

Example

Keeping with the nautical theme, we will use an example with Horseshoe crabs. These data come from Alan Agresti's book on GLMs:



We will model the number of satellites around female horseshoe crabs.

You ask - what does it mean for a crab to have satellites?

You ask - what does it mean for a crab to have satellites?

That's a bit awkward to type up.

You ask - what does it mean for a crab to have satellites?

That's a bit awkward to type up.

Let's just see online.

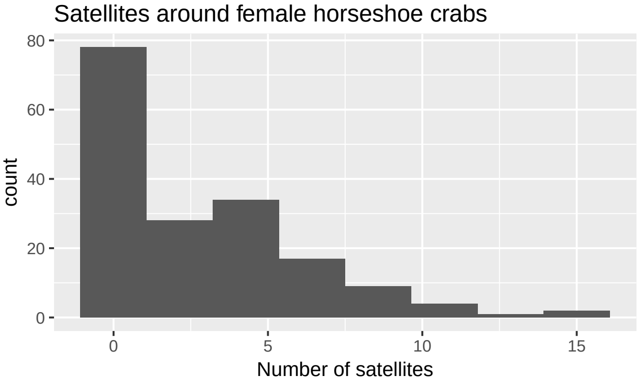
PC: http://myfwc.com/research/saltwater/crustaceans/horseshoe-crabs/facts/

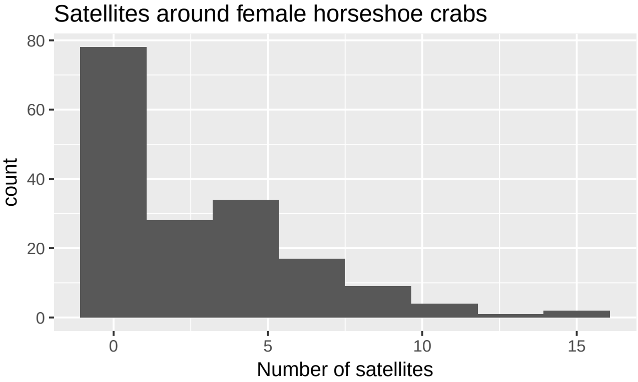
Horseshoe crab data

Load the data

```
> d <- read.table("http://www.stat.ufl.edu/~aa/glm/data/Crabs.dat",
+                 header = T)
> head(d)
  crab y weight width color spine
1    1 8   3.05  28.3     2     3
2    2 0   1.55  22.5     3     3
3    3 9   2.30  26.0     1     1
4    4 0   2.10  24.8     3     3
5    5 4   2.60  26.0     3     3
6    6 0   2.10  23.8     2     3
```

y is the number of satellites


Satellites around female horseshoe crabs
count
80
60
40
20
0
0
5
10
15
Number of satellites



The number of 0s is higher than you might expect!

Data generating process

Stochastic component:

Data generating process

Stochastic component:

$$Z_i \sim \text{Bernoulli}(p_i)$$

Data generating process

Stochastic component:

$$
\begin{aligned}
Z_i &\sim \text{Bernoulli}(p_i) \\
Y_i &\sim Z_i \text{NegBin}(\lambda_i, \theta)
\end{aligned}
$$

Data generating process

Stochastic component:

$$Z_i \sim \text{Bernoulli}(p_i)$$
$$Y_i \sim Z_i\text{NegBin}(\lambda_i, \theta)$$

Systematic component:

Data generating process

Stochastic component:

$$Z_i \sim \text{Bernoulli}(p_i)$$
$$Y_i \sim Z_i \text{NegBin}(\lambda_i, \theta)$$

Systematic component:

$$\text{logit}(p_i) = X_i \beta$$

Data generating process

Stochastic component:

$$
\begin{aligned}
Z_i &\sim \text{Bernoulli}(p_i) \\
Y_i &\sim Z_i \text{NegBin}(\lambda_i, \theta)
\end{aligned}
$$

Systematic component:

$$
\begin{aligned}
\text{logit}(p_i) &= X_i \beta \\
\log(\lambda_i) &= X_i \gamma
\end{aligned}
$$

Student: We already have the Poisson and the Negative Binomial. These each allow for some 0s. Why do we need to make this complicated mixture?

Student: We already have the Poisson and the Negative Binomial. These each allow for some 0s. Why do we need to make this complicated mixture?
Us: The mixture model allows a **more flexible** distribution of counts. It allows us to construct a data generating process that could create disproportionately more 0s than either the Poisson or Negative Binomial would have without the mixture.

Student: Is this the same as if we estimated one model for whether a crab had any satellites, and another model for the number of satellites around crabs with at least one satellite?

Student: Is this the same as if we estimated one model for whether a crab had any satellites, and another model for the number of satellites around crabs with at least one satellite?
Us: These are not quite the same, since crabs with $Z_i = 1$ may still have 0 satellites if the count portion of the mixture randomly draws a 0.
We know that crabs with satellites have $Z_i = 1$, but for those without satellites they may have $Z_i = 0$, or they may have $Z_i = 1$ and just have 0 satellites because the count drawn was 0.

Likelihood[6]

$$
\begin{aligned}
L(\beta, \gamma, \theta \mid Y) &\propto f(y \mid \beta, \gamma, \theta) \\
&= \prod_{i=1}^{n} f(y \mid \beta, \gamma\theta) \\
&= \prod_{i=1}^{n} \underbrace{\Big(f(y \mid Z = 0, \gamma) P(Z = 0 \mid \beta) + f(y \mid Z = 1, \gamma, \theta) P(Z = 1 \mid \beta) \Big)}_{\text{Apply the law of total probability}} \\
&= \prod_{i=1}^{n} \underbrace{\left(\mathbb{I}(y_i = 0)(1 - p_i) + \frac{\Gamma(\theta + y_i)}{y_i! \Gamma(\theta)} \frac{\lambda_i^{y_i} \theta^{\theta}}{(\lambda_i + \theta)^{\theta + y_i}} p_i \right)}_{\text{Substitute the PDF and PMF}} \\
&= \prod_{i=1}^{n} \underbrace{\Bigg(\mathbb{I}(y_i = 0)(1 - \text{logit}^{-1}[X_i \beta]) + \frac{\Gamma(\theta + y_i)}{y_i! \Gamma(\theta)} \frac{(\exp[X_i \gamma])^{y_i} \theta^{\theta}}{(\exp[X_i \gamma] + \theta)^{\theta + y_i}} \text{logit}^{-1}[X_i \beta] \Bigg)}_{\text{Replace parameters with inverse link of linear predictors}}
\end{aligned}
$$

[6] $\mathbb{I}(y_i = 0)$ is an *indicator function* coded 1 if $y_i = 0$ and 0 otherwise.

Code our likelihood

$$\ell(\beta, \gamma, \theta, \mid Y) = \sum_{i=1}^{n} \log \Bigg(\mathbb{I}(y_i = 0)(1 - \text{logit}^{-1}[X_i\beta]) + \frac{\Gamma(\theta + y_i)}{y_i!\Gamma(\theta)} \frac{(\exp[X_i\gamma])^{y_i}\theta^{\theta}}{(\exp[X_i\gamma] + \theta)^{\theta + y_i}} \text{logit}^{-1}[X_i\beta] \Bigg)$$

```
zinb.loglik <- function(par, y, X) {
  k <- ncol(X)
  beta <- par[1:k]
  gamma <- par[(k + 1):(2*k)]
  theta <- exp(par[(2*k + 1)])
  p <- plogis(X %*% beta)
  lambda <- exp(X %*% gamma)
  log.lik <- sum(log(
    (y == 0)*(1 - p) +
      dnbinom(y, size = theta, mu = lambda) * p
  ))
  return(log.lik)
}
```

Code our likelihood

Notes about how we coded that likelihood:

- We defined p_i and λ_i, and θ based on parameters, then coded the log likelihood as a function of those rather than as a function of β and γ directly. This is just one of several reasonable approaches.
- We used `plogis()` for the inverse logit, but we could just as well have typed `log(p / (1-p))`.
- We used `dnbinom()` rather than writing out the negative binomial density. In general, we prefer to write the density, but we used the canned version here to avoid computational issues in this particular model.

Optimize

```
opt.zinb <- optim(par = rep(0, 2*ncol(X) + 1),
                  y = y,
                  X = X,
                  fn = zinb.loglik,
                  method = "BFGS",
                  control = list(fnscale = -1),
                  hessian = TRUE)
```

Report coefficients and standard errors

```
results <- data.frame(
  Predictor = c("Intercept","Weight","Width"),
  Beta = opt.zinb$par[1:3],
  SE.Beta = sqrt(diag(-solve(opt.zinb$hessian)))[1:3],
  Gamma = opt.zinb$par[4:6],
  SE.Gamma = sqrt(diag(-solve(opt.zinb$hessian)))[4:6]
)
print(xtable(results),
      include.rownames = F)
theta <- exp(opt.zinb$par[7])
```

Report coefficients and standard errors

$$Z_i \sim \text{Bernoulli}(p_i) \qquad Y_i \sim Z_i\text{NegBin}(\lambda_i, \theta)$$

$$\text{logit}(p_i) = X_i\beta \qquad \log(\lambda_i) = X_i\gamma$$

Predictor	$\hat{\beta}$	$\widehat{SE}(\hat{\beta})$	$\hat{\gamma}$	$\widehat{SE}(\hat{\gamma})$
Intercept	-10.45	4.04	2.66	1.44
Weight	0.71	0.81	0.53	0.28
Width	0.37	0.21	-0.10	0.08

$\hat{\theta} = 5.24$

Report coefficients and standard errors

$$Z_i \sim \text{Bernoulli}(p_i) \qquad Y_i \sim Z_i\text{NegBin}(\lambda_i, \theta)$$

$$\text{logit}(p_i) = X_i\beta \qquad \log(\lambda_i) = X_i\gamma$$

Predictor	$\hat{\beta}$	$\widehat{SE}(\hat{\beta})$	$\hat{\gamma}$	$\widehat{SE}(\hat{\gamma})$
Intercept	-10.45	4.04	2.66	1.44
Weight	0.71	0.81	0.53	0.28
Width	0.37	0.21	-0.10	0.08

$\hat{\theta} = 5.24$
What do the β mean? The γ?

Simulate

```
sim.par <- mvrnorm(
  10000,
  mu = opt.zinb$par,
  Sigma = -solve(opt.zinb$hessian)
)
```

Simulate

```
sim.par <- mvrnorm(
  10000,
  mu = opt.zinb$par,
  Sigma = -solve(opt.zinb$hessian)
)
```

We found before that $\hat{\theta} = 5.24$. Can we calculate $\widehat{\mathsf{SE}}(\hat{\theta})$?

Simulate

```
sim.par <- mvrnorm(
  10000,
  mu = opt.zinb$par,
  Sigma = -solve(opt.zinb$hessian)
)
```

We found before that $\hat{\theta} = 5.24$. Can we calculate $\widehat{\text{SE}}(\hat{\theta})$?

```
> sim.theta <- exp(sim.par[,7])
> sd(sim.theta)
[1] 2.067322
```

After break: expectation maximization, missing data

After break: expectation maximization, missing data

Questions?