Poisson probability formula

$$\Pr(Y = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$

- This is the probability of exactly k (non-negative integer) occurrences of a random count variable Y.
- There is only one parameter, λ, the "arrival rate" or the Expected (average) number of occurrences, which makes the expression easy to manipulate.
- It is the limiting case of a binomial process when the number of trials n is high and the probability p is low. The sole parameter $\lambda = np$ is the expected count.
- It is not surprising that $E(Y) = \lambda$. It may be more surprising that, if it's really Poisson, also $\text{var}(Y) = \lambda$

We work with the Kiefer strike duration data recorded between 1968 and 1976. Each of the 62 observations is a strike for which the main variable of interest (dependent variable) is the duration of the strike in days. Other variables include the year of the strike and an industrial production index for the year. The logic of the model is that strikes will be shorter when industrial production is high because firms have strong incentives to settle and get back to making profits.

year	strike_duration	ip
1968	7	0.01138
1968	9	0.01138
1968	13	0.01138
1968	14	0.01138
1968	26	0.01138
1968	29	0.01138
1968	52	0.01138
1968	130	0.01138
1969	9	0.02299
1969	37	0.02299

How likely were we to see the exact sample we saw?

$$\Pr\left(Y_1 = y_1, Y_2 = y_2, \ldots Y_n = y_n | \lambda\right)$$

$$\Pr\left(Y_1 = y_1 | \lambda\right) \cdot \Pr\left(Y_2 = y_2 | \lambda\right) \cdot \cdots \cdot \Pr\left(Y_n = y_n | \lambda\right)$$

$$\frac{e^{-\lambda} \cdot \lambda^{y_1}}{y_1!} \cdot \frac{e^{-\lambda} \cdot \lambda^{y_2}}{y_2!} \cdot \cdots \cdot \frac{e^{-\lambda} \cdot \lambda^{y_n}}{y_n!}$$

Reminder the little y_i are data observed in the world.

$$\mathcal{L} = \frac{e^{-\lambda} \cdot \lambda^{y_1}}{y_1!} \cdot \frac{e^{-\lambda} \cdot \lambda^{y_2}}{y_2!} \cdot \dots \cdot \frac{e^{-\lambda} \cdot \lambda^{y_n}}{y_n!}$$

Taking log converts multiplication problems into additional problems

$$\ln \mathcal{L} = \ln \frac{e^{-\lambda} \cdot \lambda^{y_1}}{y_1!} + \ln \frac{e^{-\lambda} \cdot \lambda^{y_2}}{y_2!} + \cdots \ln \frac{e^{-\lambda} \cdot \lambda^{y_n}}{y_n!}$$

$$\ln \mathcal{L} = \ln \frac{e^{-\lambda} \cdot \lambda^{y_1}}{y_1!} + \ln \frac{e^{-\lambda} \cdot \lambda^{y_2}}{y_2!} + \cdots \ln \frac{e^{-\lambda} \cdot \lambda^{y_n}}{y_n!}$$

Zoom in on one of the elements:

$$\ln \mathcal{L}_1 = \ln \frac{e^{-\lambda} \cdot \lambda^{y_1}}{y_1!} = \ln e^{-\lambda} + \ln \lambda^{y_1} - \ln y_1! =$$

$$-\lambda + y_1 \ln \lambda - \ln(y_1) - \ln(y_1 - 1) - \ln(y_1 - 2) - \cdots - \ln 1$$

Zoom back out (there are n of these expressions to add up):

$$\ln \mathcal{L} = -n\lambda + (y_1 + y_2 + \cdots + y_n) \ln \lambda$$

$$\begin{aligned} &- \ln(y_1) - \ln(y_1 - 1) - \ln(y_1 - 2) - \cdots - \ln 1 \\ &- \ln(y_2) - \ln(y_2 - 1) - \ln(y_2 - 2) - \cdots - \ln 1 \\ &- \ln(y_n) - \ln(y_n - 1) - \ln(y_n - 2) - \cdots - \ln 1 \end{aligned}$$

The second batch of stuff changes the level of $\ln \mathcal{L}$ but does not depend on λ. Only the first batch does:

$$\ln \mathcal{L} = -n\lambda + (y_1 + y_2 + \cdots + y_n) \ln \lambda$$

$$\begin{aligned}
&- \ln(y_1) - \ln(y_1 - 1) - \ln(y_1 - 2) - \cdots - \ln 1 \\
&- \ln(y_2) - \ln(y_2 - 1) - \ln(y_2 - 2) - \cdots - \ln 1 \\
&- \ln(y_n) - \ln(y_n - 1) - \ln(y_n - 2) - \cdots - \ln 1
\end{aligned}$$

$$\max_{\lambda} \ln \mathcal{L} = -n\lambda + (y_1 + y_2 + \cdots + y_n) \ln \lambda$$

$$-n + \frac{1}{\hat{\lambda}}(y_1 + y_2 + \cdots + y_n) \equiv 0$$

$$\hat{\lambda} = (y_1 + y_2 + \cdots + y_n)/n$$

But now let's revisit that with slightly different syntax. Let $\beta = \ln \lambda$ and maximize with respect to β

$$\max_{\lambda} \ln \mathcal{L} = -n\lambda + (y_1 + y_2 + \cdots + y_n) \ln \lambda$$

$$\max_{\ln \lambda} \ln \mathcal{L} = -ne^{\ln \lambda} + (y_1 + y_2 + \cdots + y_n) \ln \lambda$$

$$\max_{\beta} \ln \mathcal{L} = -ne^{\beta} + (y_1 + y_2 + \cdots + y_n)\beta$$

$$-ne^{\hat{\beta}} + (y_1 + y_2 + \cdots + y_n) \equiv 0$$

$$e^{\hat{\beta}} = (y_1 + y_2 + \cdots + y_n)/n$$

$$\hat{\beta} = \ln ((y_1 + y_2 + \cdots + y_n)/n)$$

In the code below, `y` is the full vector of strike durations $y = 7, 9, 13, 14, 26, \ldots$, `cons` is the same length and made of 1's. The last term, which is returned by the function, is the log likelihood, for each observation $-e^{\beta} + y_i\beta$ summed over all observations, or $-ne^{\beta} + \sum y_i\beta$.

```
## log(lambda) = XB, in this case log(lambda) = B
poissonll <- function(beta) {
    y <- kiefer$strike_duration
    cons <- rep(1, length(kiefer$strike_duration))
    -sum( -exp(beta*cons) + y * beta * cons ) }
```

This function “rates” the quality of the current guess of beta by returning the (negative) log likelihood for that guess of beta. We give this function to R’s optimizer and let the optimizer find the value of beta that maximizes the log likelihood (formally, by minimizing the negative log likelihood).

```
## Use optimizer to find beta, the log lambda that minimiz
(poissonll.opt  <- optim( par=1,  poissonll ))
## Report lambda
exp(poissonll.opt$par)
```

This syntax lets us introduce covariates into the Poisson regression. Now each observation i has its own "arrival rate" λ_i, which we express logged, $\ln \lambda_i$. Instead of $\ln \lambda_i = \beta$, we let $\ln \lambda_i = \beta_1 + \beta_2 \mathrm{ip}_i$, where ip_i is the industrial production index for the year in which strike i took place.
Strike i's duration, which can be any non-negative integer, is drawn from a Poisson distribution in which the expected arrival rate is λ_i, which depends on the covariate ip_i in this particular way, $\ln \lambda_i = \beta_1 + \beta_2 \mathrm{ip}_i$.
A higher λ will tend to generate longer strikes. In particular, a unit increase in the explanatory variable will make the average strike β_2 percent longer.

Let's repeat the zoom on one of the elements with the parameterized value of λ_i

$$\ln \mathcal{L}_i = \ln \frac{e^{-\lambda_i} \cdot \lambda_i^{y_i}}{y_i!} = \ln e^{-\lambda_i} + \ln \lambda_i^{y_i} - \ln y_i! =$$

$$-\lambda_i + y_i \ln \lambda_i - \ln(y_i) - \ln(y_i - 1) - \ln(y_i - 2) - \cdots - \ln 1$$

First, rewrite this in $\ln \lambda$ syntax:

$$-e^{\ln \lambda_i} + y_i \ln \lambda_i - \ln(y_i) - \ln(y_i - 1) - \ln(y_i - 2) - \cdots - \ln 1$$

Now substitute the parameterization, $\ln \lambda_i = \beta_1 + \beta_2 \text{ip}_i$

$$-e^{\beta_1 + \beta_2 \text{ip}_i} + y_i(\beta_1 + \beta_2 \text{ip}_i) - \ln(y_i) - \ln(y_i - 1) - \ln(y_i - 2) - \cdots - \ln 1$$

Let's zoom back out and drop the extraneous terms on the right that do not depend on β or λ

$$
\begin{aligned}
\ln \mathcal{L} = \\
-e^{\beta_1 + \beta_2 \text{ip}_1} + y_1(\beta_1 + \beta_2 \text{ip}_1) \\
-e^{\beta_1 + \beta_2 \text{ip}_2} + y_2(\beta_1 + \beta_2 \text{ip}_2) \\
-e^{\beta_1 + \beta_2 \text{ip}_2} + y_3(\beta_1 + \beta_2 \text{ip}_2) \\
\cdots \\
-e^{\beta_1 + \beta_2 \text{ip}_n} + y_n(\beta_1 + \beta_2 \text{ip}_n)
\end{aligned}
$$

The maximization is now over two parameters, β_1 and β_2, instead of over one parameter β (or λ) but as before there is some $(\hat{\beta}_1, \hat{\beta}_2)$ that makes $\ln \mathcal{L}$ as large as possible. These are the MLE estimates.

Here `beta` is a two-element vector and the expression for $\ln \lambda_i$ includes the value of $\mathtt{x}_i$.

```
poissonll2 <- function(beta) {
    y <- kiefer$strike_duration
    cons <- rep(1, length(kiefer$strike_duration))
    x <- kiefer$ip
    -sum(  -exp(beta[1]*cons + beta[2] * x) + y * (beta[1]*cons + beta[2] * x) )
    }
```