

Poisson data

Properties of the MLE, Monte Carlo, and the British coal-mining disasters

Hedibert Freitas Lopes

Inspire

Poisson model

$y_i \equiv$ count of coal-mining disasters in year i , $y_i \in \{0, 1, 2, \dots\}$, $i = 1, \dots, n$.

Poisson model: $y_1, \dots, y_n \mid \lambda \stackrel{iid}{\sim} \text{Poisson}(\lambda)$,

$$\Pr(y_i \mid \lambda) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}, \quad \lambda > 0 \text{ (occurrence rate),}$$

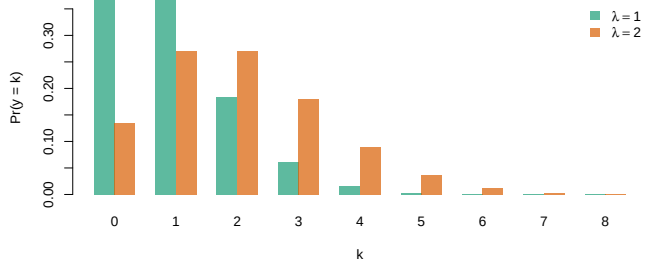
so that

$$E(y_i \mid \lambda) = V(y_i \mid \lambda) = \lambda.$$

Example

k	$\lambda = 1$	$\lambda = 2$
0	0.368	0.135
1	0.368	0.271
2	0.184	0.271
3	0.061	0.180
Sum	0.981	0.857

$\Pr(y_i = k \mid \lambda)$



Maximum likelihood estimation

$$\hat{\lambda}_{\text{MLE}} \equiv \arg \max_{\lambda > 0} p(y_1, \dots, y_n \mid \lambda)$$

$$p(y_1, \dots, y_n \mid \lambda) = \prod_{i=1}^n \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} = \frac{\lambda^{s_n} e^{-n\lambda}}{\prod_{i=1}^n y_i!}, \quad s_n = \sum_{i=1}^n y_i.$$

As a function of λ , it is usually rewritten as

$$L(\lambda \mid y_1, \dots, y_n) = c \lambda^{s_n} e^{-n\lambda}, \quad c^{-1} = \prod_{i=1}^n y_i! \text{ (constant).}$$

The likelihood depends on the data only through $(s_n, n) \Rightarrow s_n$ is a **sufficient statistic** for λ .

The MLE is the sample mean

$$\ell(\lambda) \equiv \log L(\lambda \mid y_1, \dots, y_n) = \log c + s_n \log \lambda - n\lambda$$

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = \frac{s_n}{\lambda} - n = 0 \iff \boxed{\hat{\lambda}_{\text{MLE}} = \frac{s_n}{n} = \bar{y}_n} \quad (\text{sample mean})$$

Is it a maximum?

$$\frac{\partial^2 \ell(\lambda)}{\partial \lambda^2} = -\frac{s_n}{\lambda^2} < 0 \Rightarrow \hat{\lambda}_{\text{MLE}} \text{ is a point of maximum.}$$

Also,

$$E(\hat{\lambda}_{\text{MLE}} \mid \lambda) = \frac{1}{n} \sum_{i=1}^n E(y_i \mid \lambda) = \lambda \Rightarrow \hat{\lambda}_{\text{MLE}} \text{ is unbiased.}$$

Consistency

$$V(\hat{\lambda}_{\text{MLE}} \mid \lambda) = \frac{1}{n^2} \sum_{i=1}^n V(y_i \mid \lambda) = \frac{\lambda}{n} \xrightarrow{n \rightarrow \infty} 0.$$

Unbiased and variance going to zero (so $\text{MSE} \rightarrow 0$):

$$\hat{\lambda}_{\text{MLE}} \xrightarrow{n \rightarrow \infty} \lambda \quad (\text{in probability}),$$

i.e. $\hat{\lambda}_{\text{MLE}}$ is a **consistent** estimator.

By the CLT, moreover,

$$\hat{\lambda}_{\text{MLE}} \overset{a}{\sim} N(\lambda, \lambda/n), \quad \text{SE}(\hat{\lambda}_{\text{MLE}}) = \sqrt{\hat{\lambda}_{\text{MLE}}/n}.$$

Efficiency

An unbiased estimator is **efficient** if its variance equals the Cramér–Rao lower bound:

$$\text{CRLB} = \frac{1}{I(\lambda)}, \quad I(\lambda) = n E_{y|\lambda} \left\{ \left(\frac{\partial \log p(y | \lambda)}{\partial \lambda} \right)^2 \right\} \text{ (expected Fisher information).}$$

For the Poisson, $\partial \log p(y | \lambda) / \partial \lambda = y/\lambda - 1$, so

$$\begin{aligned} I(\lambda) &= n E_{y|\lambda} \left\{ \left(\frac{y}{\lambda} - 1 \right)^2 \right\} = n \left[\frac{1}{\lambda^2} E(y^2) - \frac{2}{\lambda} E(y) + 1 \right] \\ &= n \left[\frac{\lambda+1}{\lambda} - 1 \right] = \frac{n}{\lambda} = \frac{1}{V(\hat{\lambda}_{\text{MLE}} | \lambda)}, \end{aligned}$$

using $E(y^2) = V(y) + [E(y)]^2 = \lambda(\lambda + 1)$. Hence the efficiency $e(\hat{\lambda}_{\text{MLE}}) = \{I(\lambda)\}^{-1} / V(\hat{\lambda}_{\text{MLE}} | \lambda) = 1$: **$\hat{\lambda}_{\text{MLE}}$ is efficient.**

Monte Carlo exercise

Set $\lambda = 2$, $N = 1000$ replications, and $n \in \{10, 50, 200\}$. For each n :

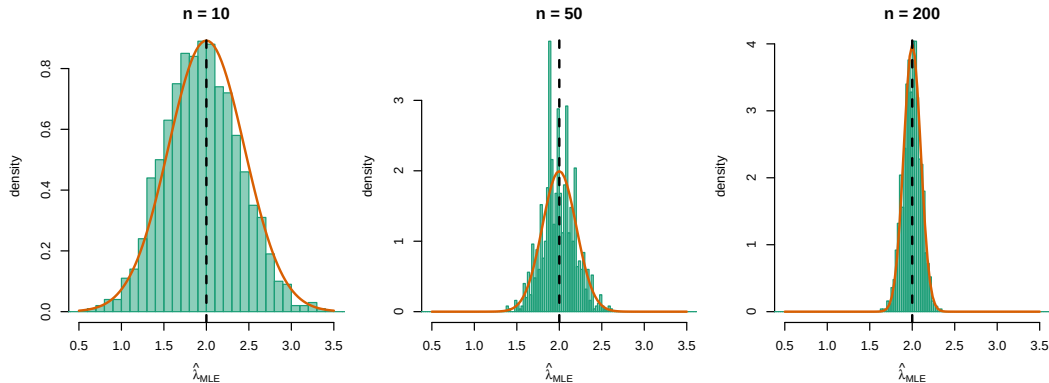
1. Draw $y_1^{(j)}, \dots, y_n^{(j)} \stackrel{iid}{\sim} \text{Poisson}(2)$, $j = 1, \dots, N$.
2. Compute $\hat{\lambda}_{\text{MLE}}^{(j)} = \bar{y}^{(j)}$.

A competitor. Since $V(y_i | \lambda) = \lambda$, the sample variance s^2 is *also unbiased* for λ , with

$$V(s^2 | \lambda) = \frac{\lambda}{n} + \frac{2\lambda^2}{n-1} > \frac{\lambda}{n} = \text{CRLB}.$$

```
y = matrix(rpois(N*n, lambda), N, n)
lhat = rowMeans(y) # MLE
s2 = apply(y, 1, var) # unbiased competitor
```

Sampling distribution of $\hat{\lambda}_{MLE}$ ($\lambda = 2$)



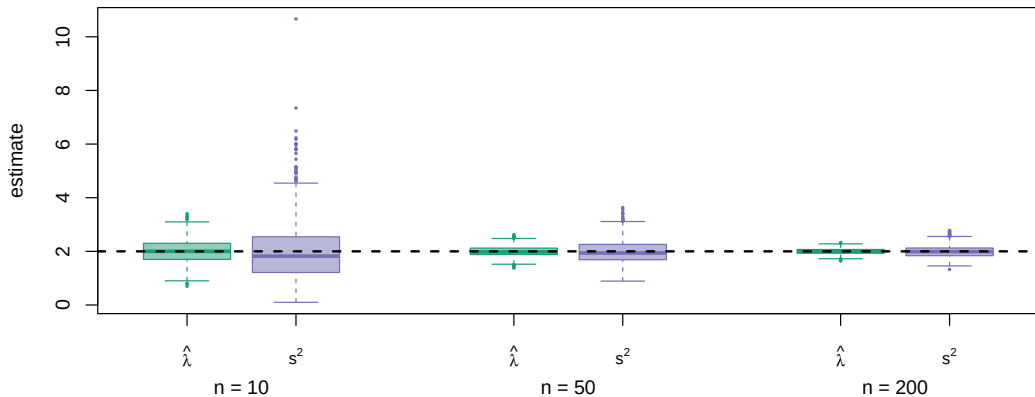
Orange: $N(\lambda, \lambda/n)$ approximation. Dashed: $\lambda = 2$. The distribution is discrete (support $\{0, 1/n, 2/n, \dots\}$) but quickly looks Gaussian.

Confirming the theory

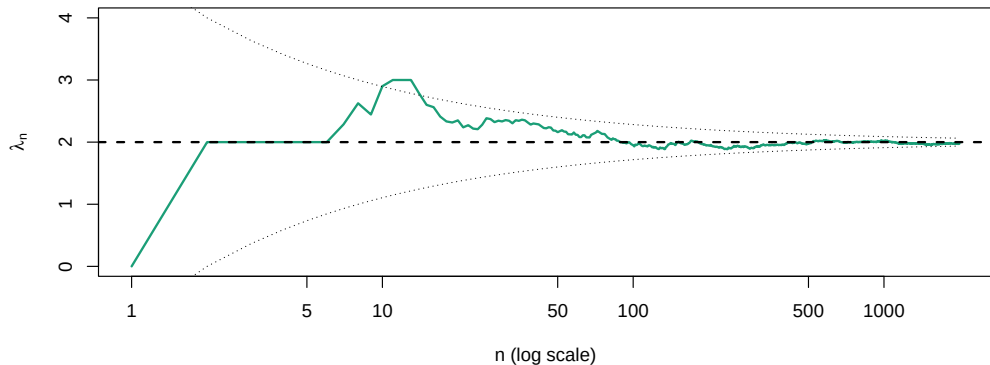
	$\hat{\lambda}_{\text{MLE}} = \bar{y}$			s^2		
	$n = 10$	$n = 50$	$n = 200$	$n = 10$	$n = 50$	$n = 200$
MC mean	2.003	1.994	1.998	2.036	1.990	1.994
MC variance	0.193	0.038	0.011	1.245	0.187	0.051
Theoretical var.	0.200	0.040	0.010	1.089	0.203	0.050
Efficiency (MC)				0.161	0.214	0.197
Efficiency (theory)		1		0.184	0.197	0.199

- ▶ Both estimators are unbiased; the MLE attains the CRLB λ/n .
- ▶ s^2 has efficiency $\approx 1/(1 + 2\lambda) = 0.2$: it needs about **5 times** more data to match $\bar{y}$.

MLE vs s^2 : boxplots

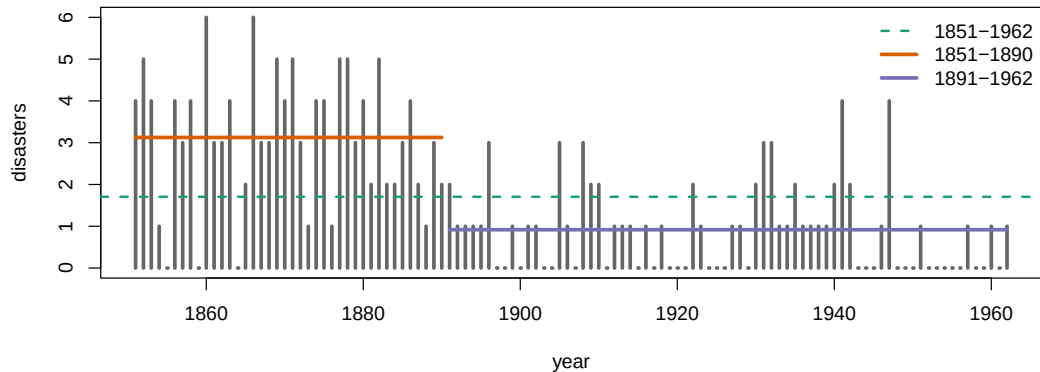


Consistency: one long sequence



- ▶ Running mean $\hat{\lambda}_n$ along 2000 Poisson(2) draws; dotted: $\lambda \pm 2\sqrt{\lambda/n}$.
- ▶ From the MC: $\Pr(|\hat{\lambda}_{\text{MLE}} - \lambda| < 0.25) \approx 0.420, 0.799, 0.984$ for $n = 10, 50, 200$.

British coal-mining disasters, 1851–1962



$n = 112$ years, $s_n = 191$ disasters. Horizontal lines: MLEs for the full period and for the two parts.

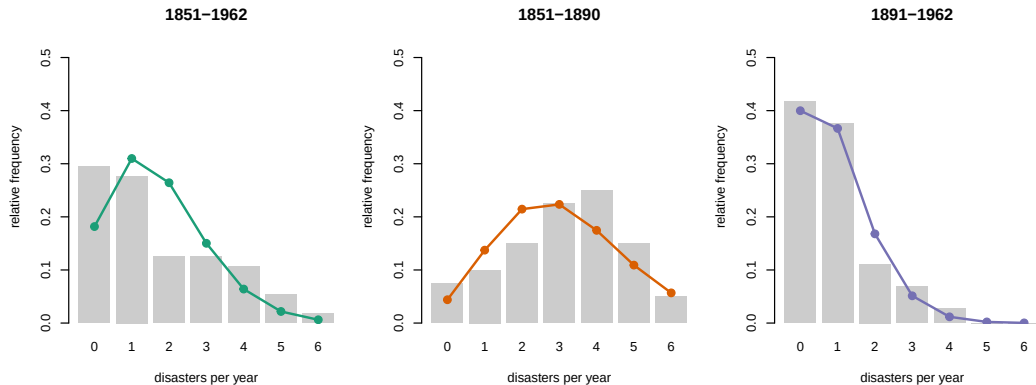
Fitting the Poisson model

	1851–1962	1851–1890	1891–1962
n	112	40	72
s_n	191	125	66
$\hat{\lambda}_{\text{MLE}} = \bar{y}$	1.705	3.125	0.917
$\text{SE} = \sqrt{\hat{\lambda}_{\text{MLE}}/n}$	0.123	0.280	0.113
95% CI	(1.464, 1.947)	(2.577, 3.673)	(0.696, 1.138)
s^2	2.714	2.574	1.063
$s^2/\bar{y}$	1.592	0.824	1.160
Dispersion test p -value	0.001	0.774	0.168

Dispersion test: $(n-1)s^2/\bar{y} \stackrel{a}{\sim} \chi^2_{n-1}$ under the Poisson model.

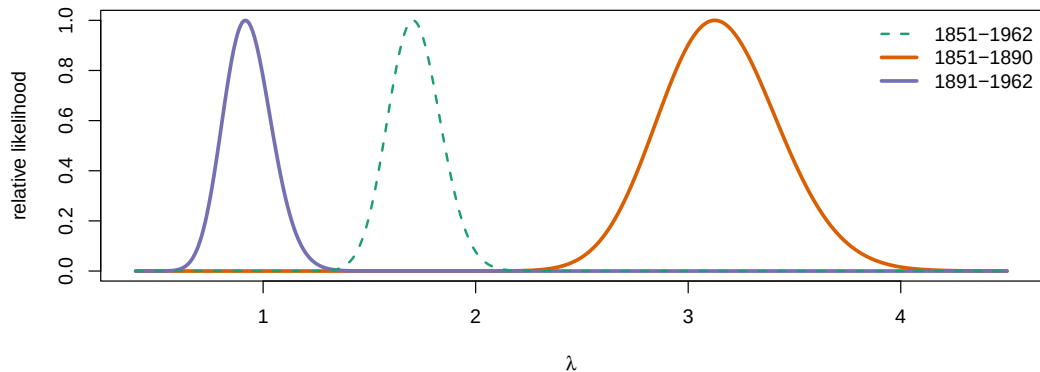
- ▶ The rate falls from about 3.1 to about 0.9 disasters/year; the two CIs are far apart.
- ▶ The full series is **overdispersed** ($s^2/\bar{y} \approx 1.6$): a single λ does not fit. Each part is consistent with Poisson.

Observed vs fitted frequencies



Grey bars: observed relative frequencies. Points: $\text{Poisson}(\hat{\lambda}_{\text{MLE}})$ probabilities. The full-period fit has too few zeros and too few large counts, the signature of a mixture of two rates.

Likelihoods for the two parts



Relative likelihoods $L(\lambda)/L(\hat{\lambda}_{\text{MLE}})$. The two parts give practically non-overlapping likelihoods; the full-period likelihood sits in between, supported by neither.

Where did the regime change?

Two-regime Poisson model with cut point τ :

$$y_t \mid \lambda_1, \lambda_2, \tau \sim \begin{cases} \text{Poisson}(\lambda_1), & t = 1851, \dots, \tau \\ \text{Poisson}(\lambda_2), & t = \tau + 1, \dots, 1962 \end{cases}$$

For each $\tau \in \{1885, \dots, 1895\}$ (5 years on each side of 1890):

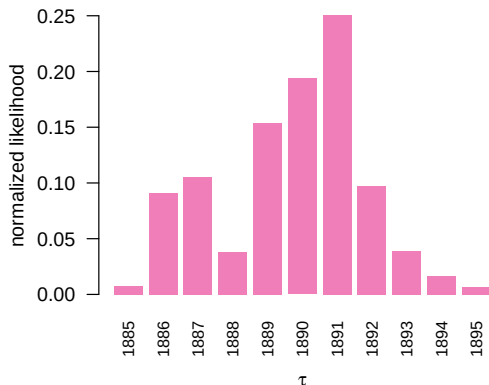
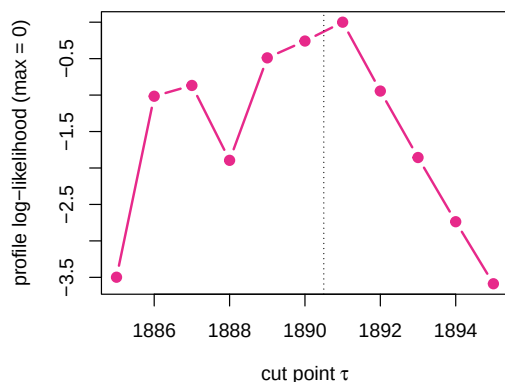
- ▶ $\hat{\lambda}_1(\tau)$ and $\hat{\lambda}_2(\tau)$ are the sample means of each part;
- ▶ profile log-likelihood $\ell_p(\tau) = s_1 \log \hat{\lambda}_1 - n_1 \hat{\lambda}_1 + s_2 \log \hat{\lambda}_2 - n_2 \hat{\lambda}_2$ (+ const.);
- ▶ normalized likelihood $w(\tau) = e^{\ell_p(\tau)} / \sum_{\tau'} e^{\ell_p(\tau')}$.

$w(\tau)$ is what a uniform prior on τ gives if (λ_1, λ_2) are profiled out (plugged in at the MLEs) rather than integrated.

Cut points around 1890

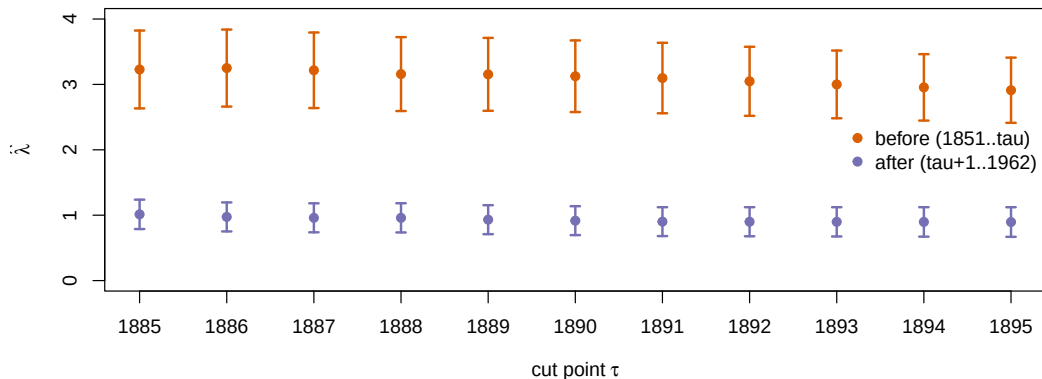
τ	n_1/n_2	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_1/\hat{\lambda}_2$	$\ell_p(\tau) - \max$	$w(\tau)$
1885	35/77	3.229	1.013	3.187	-3.50	0.008
1886	36/76	3.250	0.974	3.338	-1.02	0.091
1887	37/75	3.216	0.960	3.350	-0.87	0.105
1888	38/74	3.158	0.959	3.291	-1.89	0.038
1889	39/73	3.154	0.932	3.386	-0.49	0.154
1890	40/72	3.125	0.917	3.409	-0.26	0.194
1891	41/71	3.098	0.901	3.436	0.00	0.251
1892	42/70	3.048	0.900	3.386	-0.94	0.098
1893	43/69	3.000	0.899	3.339	-1.86	0.039
1894	44/68	2.955	0.897	3.294	-2.74	0.016
1895	45/67	2.911	0.896	3.251	-3.59	0.007

Profile likelihood of the cut point



- ▶ Maximum at $\tau = 1891$ (last year of the high-rate regime), with weight 0.251; years 1889–1892 together carry 0.696.
- ▶ Likelihood-ratio statistic vs a single λ : $2\{\ell_p(1891) - \ell(\hat{\lambda})\} = 70.0$, overwhelming evidence of a change.

Estimated rates for each cut point



95% Wald intervals. The two rates are remarkably stable across cut points: $\hat{\lambda}_1 \approx 3$ and $\hat{\lambda}_2 \approx 0.9$, i.e. the rate fell by a factor of about 3.4. The data locate the change around 1890 but cannot pin down the exact year: the “bumpy” profile reflects individual years with 0 or 3+ disasters.

Take-aways

- ▶ For iid Poisson data, $\hat{\lambda}_{\text{MLE}} = \bar{y}$ is sufficient-based, unbiased, consistent and efficient ($V = \lambda/n = \text{CRLB}$).
- ▶ Monte Carlo confirms these properties; another unbiased estimator (s^2) is only $\approx 20\%$ efficient.
- ▶ A single Poisson does not fit the coal-mining series: overdispersion is a symptom of a change in rate.
- ▶ Splitting the series gives two well-fitting Poisson models, $\hat{\lambda}_1 \approx 3.1$ and $\hat{\lambda}_2 \approx 0.9$; the profile likelihood points to a change around 1890–1891.
- ▶ Next step: a Bayesian change-point model, $p(\tau, \lambda_1, \lambda_2 \mid y)$, integrating rather than profiling the rates.