Inference about a Population Rate (λ)

JH notes on rates

Sahir Bhatnagar and James Hanley

EPIB 607
Department of Epidemiology, Biostatistics, and Occupational Health
McGill University

sahir.bhatnagar@mcgill.ca
https://sahirbhatnagar.com/EPIB607/

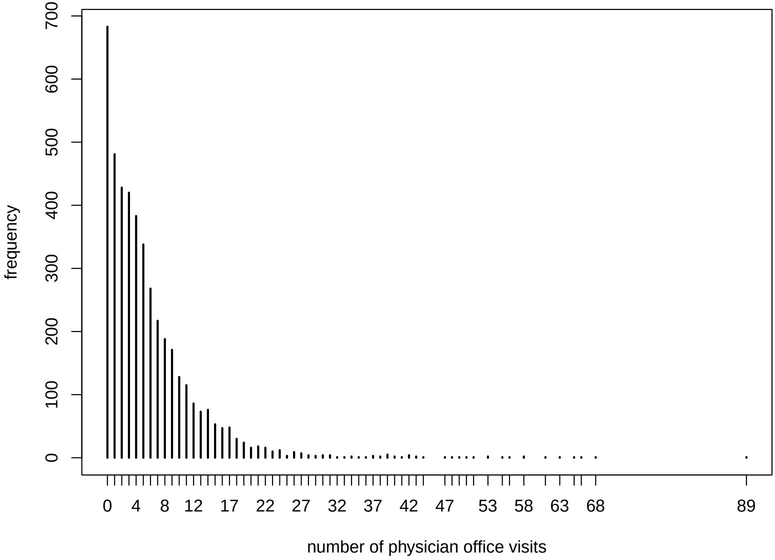
October 24, 2018

Poisson Model for Sampling Variability of a Count in a Given Amount of "Experience"

Motivating example: Demand for medical care

- Data from the US National Medical Expenditure Survey (NMES) for 1987/88
- 4406 individuals, aged 66 and over, who are covered by Medicare, a public insurance program
- The objective of the study was to model the demand for medical care - as captured by <u>the number</u> of physician/non-physician office and hospital outpatient visits - by the covariates available for the patients.

Motivating example: Demand for medical care



Some observations about the previous plot

- Discrete outcome $\rightarrow$ 1, 2, 3, … visits

Some observations about the previous plot

- Discrete outcome $\rightarrow$ 1, 2, 3, … visits
- There are rare events, e.g. 1 individual with 89 visits

Some observations about the previous plot

- Discrete outcome $\rightarrow$ 1, 2, 3, ... visits
- There are rare events, e.g. 1 individual with 89 visits
- The data are far from normally distributed

Some observations about the previous plot

- Discrete outcome $\rightarrow$ 1, 2, 3, ... visits
- There are rare events, e.g. 1 individual with 89 visits
- The data are far from normally distributed
- Can theoretically go on forever

The Poisson Distribution

- The binomial distribution was derived by starting with an experiment consisting of trials or draws and applying the laws of probability to various outcomes of the experiment.

The Poisson Distribution

- The binomial distribution was derived by starting with an experiment consisting of trials or draws and applying the laws of probability to various outcomes of the experiment.
- There is no simple experiment on which the Poisson distribution is based, although we will shortly describe how it can be obtained by certain limiting operations.

The Poisson Distribution: what it is, and features

- The (infinite number of) probabilities $P_0, P_1, ..., P_y, ...,$ of observing $Y = 0, 1, 2, \ldots, y, \ldots$ events in a given amount of "experience."

The Poisson Distribution: what it is, and features

- The (infinite number of) probabilities $P_0, P_1, ..., P_y, ...,$ of observing $Y = 0, 1, 2, \ldots, y, \ldots$ events in a given amount of "experience."

- These probabilities, $P(Y = k) \rightarrow$ `dpois()`, are governed by a single parameter, the mean $E[Y] = \mu$ which represents the expected **number** of events in the amount of experience actually studied.

The Poisson Distribution: what it is, and features

- The (infinite number of) probabilities $P_0, P_1, ..., P_y, ...,$ of observing $Y = 0, 1, 2, \ldots, y, \ldots$ events in a given amount of "experience."

- These probabilities, $P(Y = k) \rightarrow$ `dpois()`, are governed by a single parameter, the mean $E[Y] = \mu$ which represents the expected **number** of events in the amount of experience actually studied.

- We say that a random variable $Y \sim \mathrm{Poisson}(\mu)$ distribution if

$$P(Y = k) = \frac{\mu^k}{k!} e^{-\mu}, \quad k = 0, 1, 2, \ldots$$

The Poisson Distribution: what it is, and features

- The (infinite number of) probabilities $P_0, P_1, ..., P_y, ...,$ of observing $Y = 0, 1, 2, \ldots, y, \ldots$ events in a given amount of "experience."

- These probabilities, $P(Y = k) \rightarrow$ `dpois()`, are governed by a single parameter, the mean $E[Y] = \mu$ which represents the expected **number** of events in the amount of experience actually studied.

- We say that a random variable $Y \sim \text{Poisson}(\mu)$ distribution if

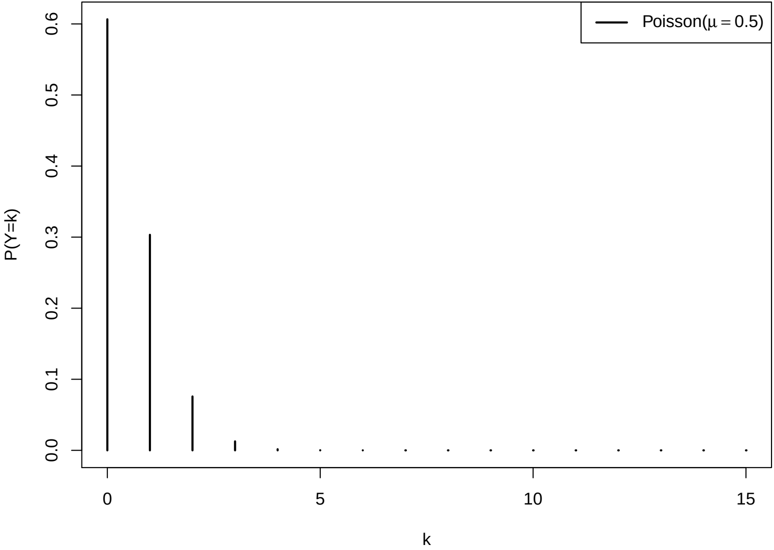
$$P(Y = k) = \frac{\mu^k}{k!} e^{-\mu}, \quad k = 0, 1, 2, \ldots$$

- Note: in `dpois()` μ is referred to as `lambda`

- Note the distinction between μ and λ
  - μ: expected **number** of events
  - λ: **rate** parameter

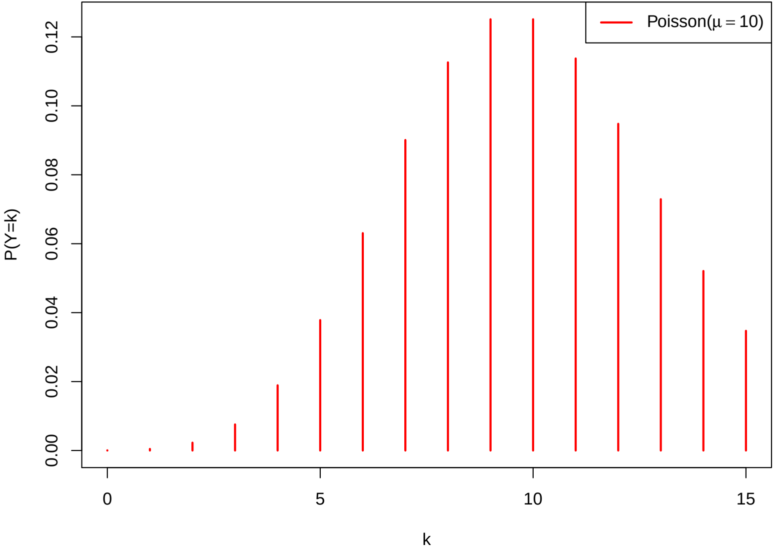
The probability mass function for $\mu = 0.5$

```r
dpois(x = 0:15, lambda = 0.5)
```

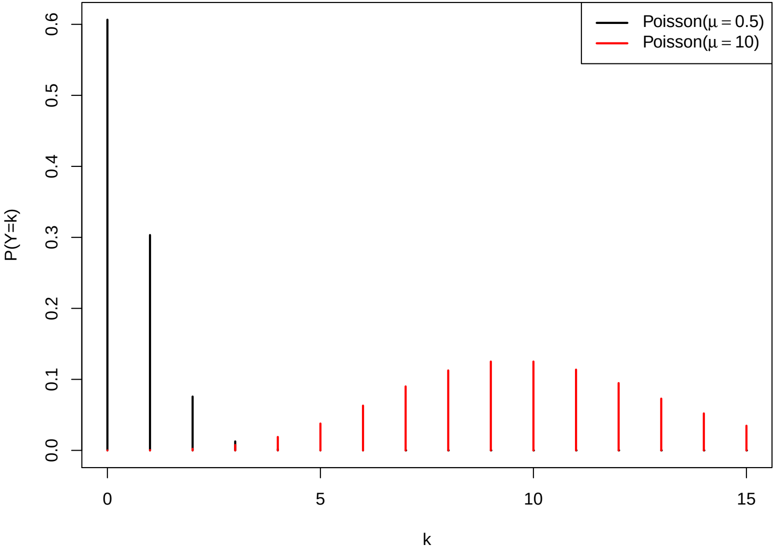


The probability mass function for $\mu = 10$

```r
dpois(x = 0:15, lambda = 10)
```



The probability mass function



The Poisson Distribution: what it is, and features

- $\sigma_Y^2 = \mu \ \rightarrow \ \ \sigma_Y = \sqrt{\mu}$.

The Poisson Distribution: what it is, and features

- $\sigma_Y^2 = \mu \ \rightarrow \ \ \sigma_Y = \sqrt{\mu}$.
- Approximated by $\mathcal{N}(\mu, \sqrt{\mu})$ when $\mu >> 10$

The Poisson Distribution: what it is, and features

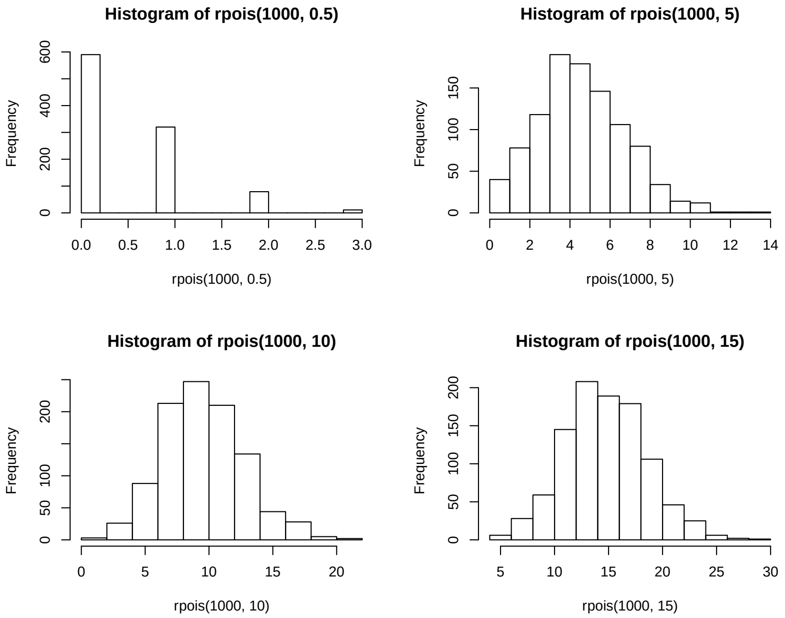
- $\sigma_Y^2 = \mu \;\rightarrow\; \sigma_Y = \sqrt{\mu}$.

- Approximated by $\mathcal{N}(\mu, \sqrt{\mu})$ when $\mu >> 10$

- Open-ended (unlike Binomial), but in practice, has finite range.

- Poisson data sometimes called ”numerator only”: (unlike Binomial) may not “see” or count “non-events”

Normal approximation to Poisson is the CLT in action



How it arises

- Count of events or items that occur randomly, with low homogeneous intensity, in time, space, or ‘item’-time (e.g. person–time).

How it arises

- Count of events or items that occur randomly, with low homogeneous intensity, in time, space, or 'item'-time (e.g. person–time).
- Binomial(n, π) when $n \to \infty$ and $\pi \to 0$, but $n \times \pi = \mu$ is finite.

How it arises

- Count of events or items that occur randomly, with low homogeneous intensity, in time, space, or 'item'-time (e.g. person–time).
- Binomial(n, π) when $n \to \infty$ and $\pi \to 0$, but $n \times \pi = \mu$ is finite.
- $Y \sim Poisson(\mu_Y)$ if time (T) between events follows an $T \sim \text{Exponential}(\mu_T = 1/\mu_Y)$. http://www.epi.mcgill.ca/hanley/bios601/Intensity-Rate/Randomness_poisson.pdf

How it arises

- Count of events or items that occur randomly, with low homogeneous intensity, in time, space, or 'item'-time (e.g. person–time).

- Binomial(n, π) when $n \to \infty$ and $\pi \to 0$, but $n \times \pi = \mu$ is finite.

- $Y \sim Poisson(\mu_Y)$ if time (T) between events follows an $T \sim \text{Exponential}(\mu_T = 1/\mu_Y)$. http://www.epi.mcgill.ca/hanley/bios601/Intensity-Rate/Randomness_poisson.pdf

- As sum of ≥ 2 *independent* Poisson random variables, with same **or different** μ's: $Y_1 \sim \text{Poisson}(\mu_1)$ $Y_2 \sim \text{Poisson}(\mu_2) \Rightarrow Y = Y_1 + Y_2 \sim \text{Poisson}(\mu_1 + \mu_2)$.

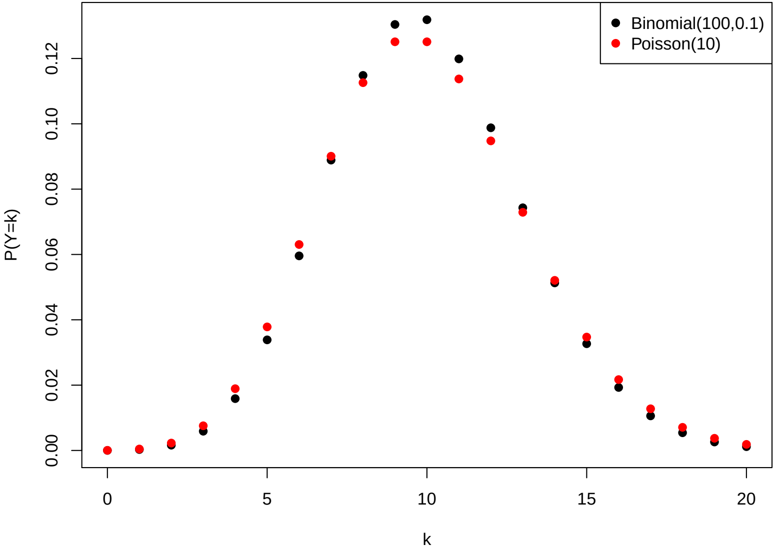
Poisson distribution as a limit

The rationale for using the Poisson distribution in many situations is provided by the following proposition.

Proposition 1 (Limit of a binomial is Poisson)

Suppose that $Y \sim Binomial(n, \pi)$. If we let $\pi = \mu/n$, then as $n \to \infty$, $Binomial(n, \pi) \to Poisson(\mu)$. Another way of saying this: for large n and small π, we can approximate the $Binomial(n, \pi)$ probability by the $Poisson(\mu = n\pi)$.

Poisson approximation to the Binomial



Examples

- numbers of asbestos fibres
- deaths from horse kicks*
- needle-stick or other percutaneous injuries
- bus-driver accidents*
- twin-pairs*
- radioactive disintegrations*
- flying-bomb hits*
- white blood cells
- typographical errors
- cell occupants – in a given volume, area, line-length, population-time, time, etc. [1]

[1]* included in http://www.epi.mcgill.ca/hanley/bios601/Intensity-Rate/

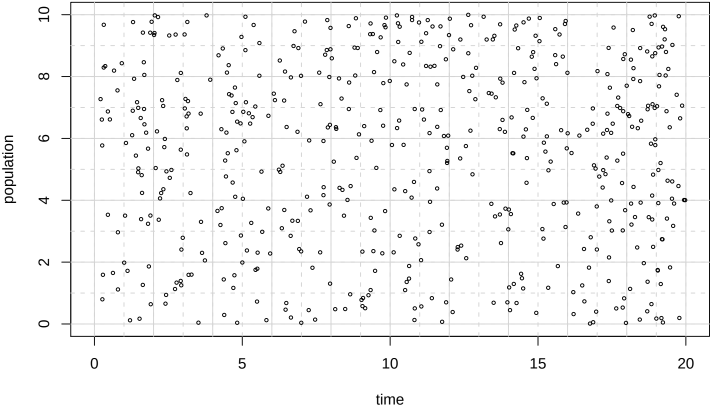


Fig.: Events in Population-Time randomly generated from intensities that are constant within (2 squares high by 2 squares wide) ‘panels’, but vary between such panels. In Epidemiology, each square might represent a number of units of population-time, and each dot an event.

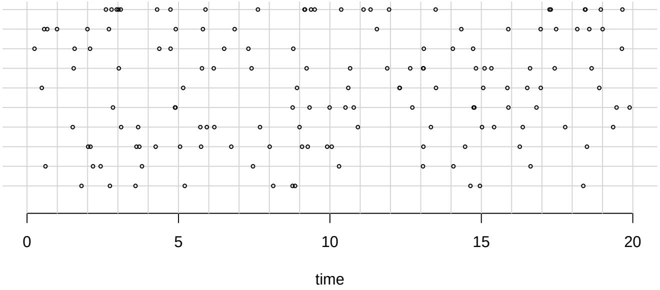


Fig.: Events in Time: 10 examples, randomly generated from constant over time intensities. Simulated with 1000 Bernoulli($_{\text{small}}\ \pi$)'s per time unit.

Does the Poisson Distribution apply to.. ?

1. Yearly variations in numbers of persons killed in plane crashes
2. Daily variations in numbers of births
3. Weekly variations in numbers of births
4. Daily variations in numbers of deaths
5. Daily variations in numbers of traffic accidents
6. Variations across cookies/pizzas in numbers of chocolate chips/olives

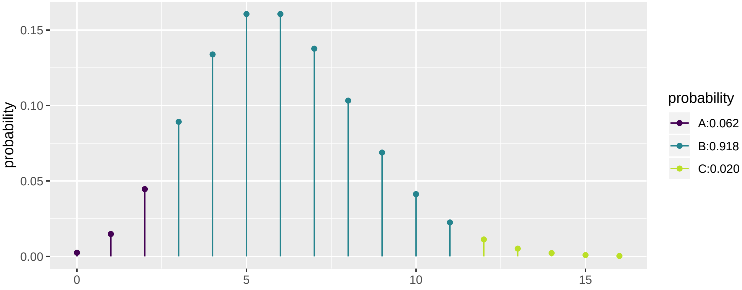
Inference regarding μ, based on observed count y

Confidence interval for μ

- If the CLT hasn't kicked in, then the usual CI might not be appropriate:

$$\text{point-estimate} \pm t^{\star} * \text{standard error}$$

```
mosaic::xqpois(c(0.025, 0.975), lambda = 6)
```




```
## [1]  2 11
```

Confidence interval for μ

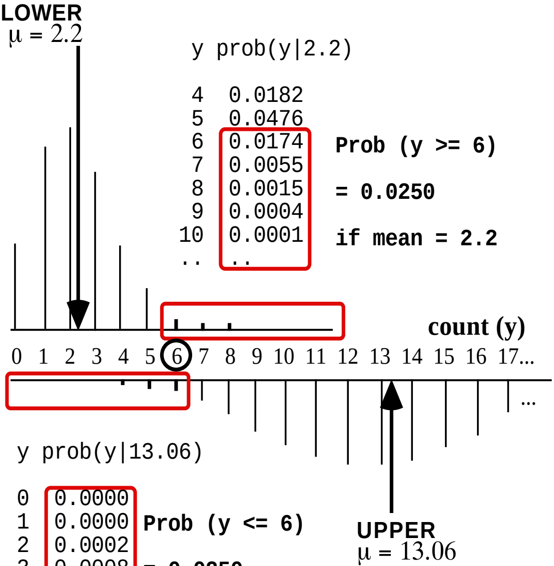
```r
manipulate::manipulate(
mosaic::xqpois(c(0.025, 0.975), lambda = LAMBDA),
LAMBDA = manipulate::slider(1, 200, step = 1))
```

Confidence interval for μ

- Similar to the binomial (Clopper-Pearson CI), we consider a *first-principles* $100(1-\alpha)\%$ CI $[\mu_{LOWER},\ \mu_{UPPER}]$ such that

$$P(Y \geq y \mid \mu_{LOWER}) = \alpha/2 \quad \text{and} \quad P(Y \leq y \mid \mu_{UPPER}) = \alpha/2.$$

- For example, the 95% CI for μ, based on $y = 6$, is $[\underline{2.20}, \underline{13.06}]$.



y	prob(y|2.2)
4	0.0182
5	0.0476
6	0.0174
7	0.0055
8	0.0015
9	0.0004
10	0.0001
..	..

Prob (y >= 6)

= 0.0250

if mean = 2.2

y	prob(y|13.06)
0	0.0000
1	0.0000
2	0.0002
3	0.0008
4	0.0026
5	0.0067
6	0.0147
7	0.0274
..	..

Prob (y <= 6)

= 0.0250

if mean = 13.06

6 **observed count**

Poisson 95% CI for μ when $y = 6$

```
# upper limit --> lower tail needs 2.5%
manipulate::manipulate(
mosaic::xppois(6, lambda = LAMBDA),
LAMBDA = manipulate::slider(0.01, 20, step = 0.01))

# lower limit --> upper tail needs 2.5%
# when lower.tail=FALSE, ppois doesnt include k, i.e., P(Y > k)
manipulate::manipulate(
mosaic::xppois(5, lambda = LAMBDA, lower.tail = FALSE),
LAMBDA = manipulate::slider(0.01, 20, step = 0.01))
```

Confidence interval for μ

- For a given confidence level, there is one CI for each value of *y*.
- Each one can be worked out by trial and error, or – as has been done for the last 80 years – directly from the (exact) link between the tail areas of the Poisson and **Gamma** distributions.
- These CI's – for *y* up to at least 30 – were found in special books of statistical tables or in textbooks.
- As you can check, *z*-based intervals are more than adequate beyond this *y*. **Today**, if you have access to R (or `Stata` or SAS) you can obtain the first principles CIs directly **for *any* value of** *y*.

80%, 90% and 95% CI for mean count μ if we observe 0 to 30 events in a certain amount of experience

y	95%		90%		80%	
0	0.00	3.69	0.00	3.00	0.00	2.30
1	0.03	5.57	0.05	4.74	0.11	3.89
2	0.24	7.22	0.36	6.30	0.53	5.32
3	0.62	8.77	0.82	7.75	1.10	6.68
4	1.09	10.24	1.37	9.15	1.74	7.99
5	1.62	11.67	1.97	10.51	2.43	9.27
6	2.20	13.06	2.61	11.84	3.15	10.53
7	2.81	14.42	3.29	13.15	3.89	11.77
8	3.45	15.76	3.98	14.43	4.66	12.99
9	4.12	17.08	4.70	15.71	5.43	14.21
10	4.80	18.39	5.43	16.96	6.22	15.41
11	5.49	19.68	6.17	18.21	7.02	16.60
12	6.20	20.96	6.92	19.44	7.83	17.78
13	6.92	22.23	7.69	20.67	8.65	18.96
14	7.65	23.49	8.46	21.89	9.47	20.13
15	8.40	24.74	9.25	23.10	10.30	21.29
16	9.15	25.98	10.04	24.30	11.14	22.45
17	9.90	27.22	10.83	25.50	11.98	23.61
18	10.67	28.45	11.63	26.69	12.82	24.76
19	11.44	29.67	12.44	27.88	13.67	25.90
20	12.22	30.89	13.25	29.06	14.53	27.05
21	13.00	32.10	14.07	30.24	15.38	28.18
22	13.79	33.31	14.89	31.41	16.24	29.32
23	14.58	34.51	15.72	32.59	17.11	30.45
24	15.38	35.71	16.55	33.75	17.97	31.58

95% CI for mean count μ with q function

- To obtain these in R we use the natural link between the Poisson and the *gamma* distributions.[2]
- In R, e.g., the 95% limits for μ based on $y = 6$ are obtained as

```
qgamma(p = c(0.025,0.975), shape = c(6, 7))
```

```
## [1]  2.201894 13.059474
```

- More generically, for *any* y, as

```
qgamma(p = c(0.025,0.975), shape = c(y, y+1))
```

[2] details found here

95% CI for mean count μ with canned function

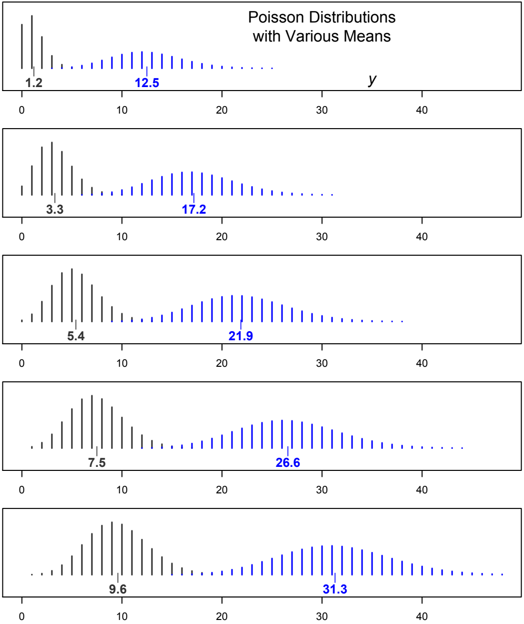
- These limits can <u>also</u> be found using the canned function in R

```
stats::poisson.test(6)
```

```
##
## ^^IExact Poisson test
##
## data:  6 time base: 1
## number of events = 6, time base = 1, p-value = 0.0005942
## alternative hypothesis: true event rate is not equal to 1
## 95 percent confidence interval:
##   2.201894 13.059474
## sample estimates:
## event rate
##          6
```

z-based confidence intervals

once μ is in the upper teens, the Poisson $\rightarrow$ the Normal



z-based confidence intervals

- Thus, a plus/minus CI based on SE = $\hat{\sigma} = \sqrt{\hat{\mu}} = \sqrt{y}$, is simply
$$[\mu_L,\ \mu_U] = y\ \pm\ z^{\star} \times \sqrt{y}.$$

- Equivalently we can use the **q** function:

$$qnorm(p = c(0.025, 0.975), mean = y, sd = \sqrt{y})$$

z-based confidence intervals

- Thus, a plus/minus CI based on SE = $\hat{\sigma} = \sqrt{\hat{\mu}} = \sqrt{y}$, is simply
$$[\mu_L,\ \mu_U] = y\ \pm\ z^{\star} \times \sqrt{y}.$$

- Equivalently we can use the **q** function:
$$qnorm(p = c(0.025, 0.975), mean = y, sd = \sqrt{y})$$

- From a single realization y of a $N(\mu, \sigma_Y)$ random variable, we can't estimate **both** μ and σ_Y: for a SE, we would have to use *outside* information on σ_Y.

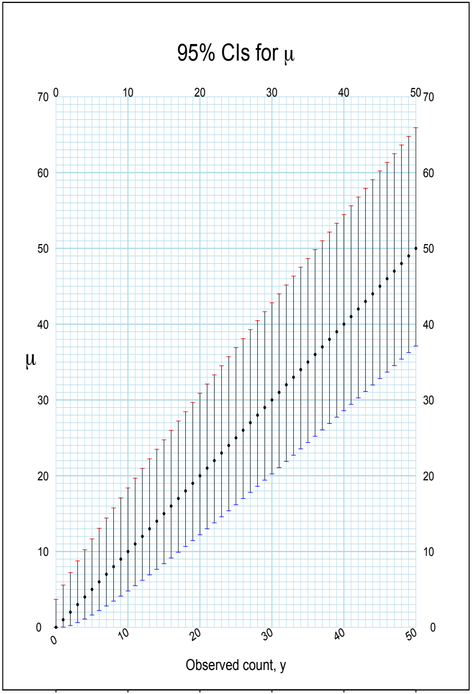
z-based confidence intervals

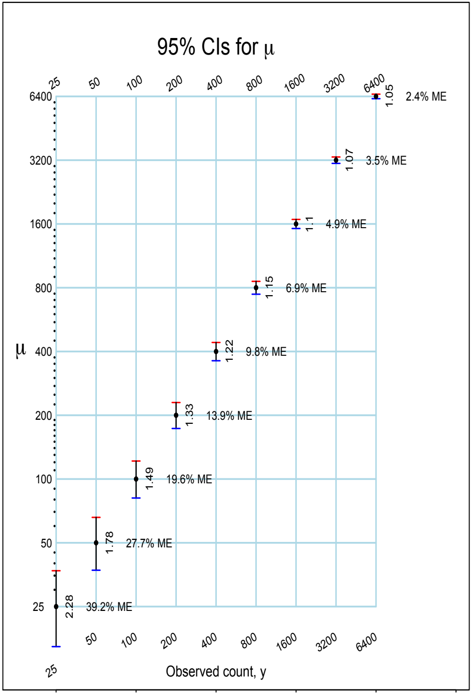
- Thus, a plus/minus CI based on SE = $\hat{\sigma} = \sqrt{\hat{\mu}} = \sqrt{y}$, is simply
$$[\mu_L,\ \mu_U] = y\ \pm\ z^{\star} \times \sqrt{y}.$$

- Equivalently we can use the **q** function:
$$qnorm(p = c(0.025, 0.975), mean = y, sd = \sqrt{y})$$

- From a single realization y of a $N(\mu, \sigma_Y)$ random variable, we can't estimate **both** μ and σ_Y: for a SE, we would have to use *outside* information on σ_Y.

- In the Poisson(μ) distribution, $\sigma_Y = \sqrt{\mu}$, so we calculate a "<u>model-based</u>" SE.


95% CIs for μ
μ
Observed count, y
0
10
20
30
40
50
60
70


95% CIs for μ
μ
Observed count, y
25
50
100
200
400
800
1600
3200
6400
2.28
39.2% ME
1.78
27.7% ME
1.49
19.6% ME
1.33
13.9% ME
1.22
9.8% ME
1.15
6.9% ME
1.1
4.9% ME
1.07
3.5% ME
1.05
2.4% ME

Inference regarding an event rate parameter λ, based on observed number of events y in a known amount of population-time (PT)

Rates are better for comparisons

- So far, we have focused on inference regarding μ, the expected **number** of events in the amount of experience actually studied.

Rates are better for comparisons

- So far, we have focused on inference regarding μ, the expected **number** of events in the amount of experience actually studied.
- However, for comparison purposes, the frequency is more often expressed as a **rate**, **intensity** or **incidence density (ID)**.

Rates are better for comparisons

- So far, we have focused on inference regarding μ, the expected **number** of events in the amount of experience actually studied.
- However, for comparison purposes, the frequency is more often expressed as a **rate**, **intensity** or **incidence density (ID)**.
- Example: we convert $y = 211$ deaths from lung cancer in 232978 women-years (WY) in the age-group 55-60 in Quebec in 2002 into a rate or incidence density of 211/(232,978WY) = 0.00091/WY or **91** per 100,000WY.

Rates are better for comparisons

- So far, we have focused on inference regarding μ, the expected **number** of events in the amount of experience actually studied.

- However, for comparison purposes, the frequency is more often expressed as a **rate**, **intensity** or **incidence density (ID)**.

- Example: we convert $y = 211$ deaths from lung cancer in 232978 women-years (WY) in the age-group 55-60 in Quebec in 2002 into a rate or incidence density of 211/(232,978WY) = 0.00091/WY or **91** per 100,000WY.

- This makes it easier to compare with the rate in the age-group 55-60 in Quebec in 1971, namely 33 lung cancer deaths in 131200WY, or 0.00025/WY = **25** per 100,000WY.

Rates are better for comparisons

- The *statistic*, the empirical rate or empirical incidence density, is
$$rate = \hat{ID} = \hat{\lambda} = y/\mathrm{PT}.$$

- where y is the observed number of events and PT is the amount of Population-Time in which these events were observed.

- We think of $\hat{ID}$ or $\hat{\lambda}$ as a point estimate of the (theoretical) Incidence Density *parameter*, ID or λ.

CI for the rate parameter λ

- To calculate a CI for the ID parameter, we **treat the PT denominator as a constant**, and the **numerator, y, as a Poisson random variable**, with expectation $E[y] = \mu = \lambda \times PT$, so that

$$\lambda = \mu \div \text{PT},$$

$$\hat{\lambda} = \hat{\mu} \div \text{PT} = y \div \text{ PT},$$

$$\boxed{\text{CI for } \lambda = \{\text{CI for } \mu\} \div \text{PT.}} \qquad (1)$$

CI for the rate parameter λ

- $y = 211$ deaths from lung cancer in 2002 leads to a 95% CI for μ:

```r
qgamma(p = c(0.025, 0.975), shape = c(211, 212))
```

```
## [1] 183.4885 241.4725
```

CI for the rate parameter λ

- $y = 211$ deaths from lung cancer in 2002 leads to a 95% CI for μ:

```
qgamma(p = c(0.025, 0.975), shape = c(211, 212))
```

```
## [1] 183.4885 241.4725
```

- From this we can calculate the 95% CI per 100,000 WY for λ using a PT=232978 years:

```
qgamma(p = c(0.025, 0.975), shape = c(211, 212)) / 232978 * 1e5
```

```
## [1]  78.75788 103.64607
```

CI for the rate parameter λ

- $y = 211$ deaths from lung cancer in 2002 leads to a 95% CI for μ:

```
qgamma(p = c(0.025, 0.975), shape = c(211, 212))
```

```
## [1] 183.4885 241.4725
```

- From this we can calculate the 95% CI per 100,000 WY for λ using a PT=232978 years:

```
qgamma(p = c(0.025, 0.975), shape = c(211, 212)) / 232978 * 1e5
```

```
## [1]  78.75788 103.64607
```

- $y = 33$ deaths from lung cancer in 131200 women-years in 1971 leads to a 95% CI per 100,000 WY for λ of

```
qgamma(c(0.025,0.975), c(33,34)) / 131200 * 1e5
```

```
## [1] 17.31378 35.32338
```

CI for the rate parameter λ using canned function

```
stats::poisson.test(x = 33, T = 131200)
```

```
##
## ^^IExact Poisson test
##
## data:  33 time base: 131200
## number of events = 33, time base = 131200, p-value < 2.2e-16
## alternative hypothesis: true event rate is not equal to 1
## 95 percent confidence interval:
##  0.0001731378 0.0003532338
## sample estimates:
##   event rate
## 0.0002515244
```

$$\text{Test of } H_0 : \mu = \mu_0 \quad \Leftrightarrow \quad \lambda = \lambda_0$$

Statistical evidence and the p-value

Recall:

- P-Value = Prob[y or more extreme | H_0]
- With 'more extreme' determined by whether H_{alt} is 1-sided or 2-sided.
- For a **formal test**, at level α, compare this P-value with α.