

The Poisson Random Variable describes the number of occurrences of rare events in a period of time or measure of area. In this study, we use the suggestion of Devore's Fifth edition Problem 26 of § 7.2, to use the asymptotic normality to find a confidence interval for the Poisson parameter λ .

For the rate constant $\lambda > 0$, the Poisson pmf is defined by the formula

$$p(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad \text{for } x = 0, 1, 2, \dots$$

Since $E(X) = V(X) = \lambda$, an estimator for λ and σ^2 is the sample mean

$$\bar{X} = S^2 = \frac{1}{n} \sum_{i=1}^n X_i.$$

The sample mean is the maximum likelihood estimator $\hat{\lambda}$.

Now, the standardized sample mean is approximately normal for large sample sizes

$$Z \sim \frac{\bar{X} - \lambda}{\sqrt{\lambda/n}}$$

Thus, for confidence level α and critical value $\Phi(z_{\alpha/2}) = 1 - \alpha/2$, we are confident that the random variable falls in the range

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha.$$

Following the derivation of the Agresti-Coull CI for a proportion, we solve for λ in terms of \bar{X} in the inequality. Thus we are $1 - \alpha$ confident that

$$\bar{X} + \frac{z_{\alpha/2}^2}{2n} - z_{\alpha/2} \sqrt{\frac{\bar{X}}{n} + \frac{z_{\alpha/2}^2}{4n^2}} < \lambda < \bar{X} + \frac{z_{\alpha/2}^2}{2n} + z_{\alpha/2} \sqrt{\frac{\bar{X}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}. \quad (1)$$

We apply this CI to two data sets. According to Bulmer, *Principles of Statistics*, Dover, 1979, Poisson developed his distribution to account for decisions of juries. But it was not noticed until von Bortkiewicz's book *Das Gesetz der kleinen Zahlen*, 1898, which applied it to the occurrences of rare events. We use one of his data sets: the number of deaths from horse kicks in the Prussian army, which is one of the classic examples. The counts give the number of deaths per army corps per year during 1875–1894.

The other is Devore's Problem 14.16. It is from the article "On the nature of Sister-Chromatid Exchanges in 5-Bromodeoxyuridine-Substituted Chromosomes" that appeared in *Genetics*, 1979.

Finally, we compute the coverage probabilities of this interval. This computation is similar to the coverage probabilities we found for the traditional vs. the Agresti-Coull intervals for the proportion. For a given λ , we find the probability that the interval (1) captures it. We need to know that the sum of independent Poisson variables is Poisson. Indeed, if $X \sim \text{Pois}(\lambda)$ and

$Y \sim \text{Pois}(\kappa)$ then using the binomial formula, $W = X + Y$ has the pdf

$$\begin{aligned} P(W = w) &= P(X + Y = w) \\ &= P((X, Y) \in \{(0, w), (1, w - 1), \dots, (w, 0)\}) \\ &= \sum_{x=0}^w P(X = x \text{ and } Y = w - x) \\ &= \sum_{x=0}^w P(X = x)P(Y = w - x) \\ &= \sum_{x=0}^w \frac{e^{-\lambda} \lambda^x}{x!} \cdot \frac{e^{-\kappa} \kappa^{w-x}}{(w-x)!} \\ &= \frac{e^{-\lambda-\kappa}}{w!} \sum_{x=0}^w \frac{w!}{x!(w-x)!} \lambda^x \kappa^{w-x} \\ &= \frac{e^{-\lambda-\kappa} (\lambda + \kappa)^w}{w!}. \end{aligned}$$

It follows by induction that if $X_i \sim \text{Pois}(\lambda_i)$ for $i = 1, 2, \dots, n$ are random variables which are independent, then

$$X_1 + X_2 + \dots + X_n \sim \text{Pois}(\lambda_1 + \lambda_2 + \dots + \lambda_n).$$

Thus given λ , for a random sample of size n taken from $\text{Pois}(\lambda)$, the probability, P , that $-z_{\alpha/2} < Z < z_{\alpha/2}$ is

$$n\lambda - z_{\alpha/2}\sqrt{n\lambda} < X_1 + X_2 + \dots + X_n < n\lambda + z_{\alpha/2}\sqrt{n\lambda}$$

and is given in terms of the Poisson cdf, $\text{ppois}(x; n\lambda) = P(X_1 + X_2 + \dots + X_n \leq x)$ is for almost all λ ,

$$P = \text{ppois}(n\lambda + z_{\alpha/2}\sqrt{n\lambda}, n\lambda) - \text{ppois}(n\lambda - z_{\alpha/2}\sqrt{n\lambda}, n\lambda).$$

Because the Poisson rv is discrete, this quantity is discontinuous in λ . At the jumps, we have not taken care to include or exclude endpoints correctly, but this happens at most at countably many λ for each n . We have plotted this function for $n = 10$ and $n = 100$.

R Session:

R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

[R.app GUI 1.31 (5538) powerpc-apple-darwin8.11.1]

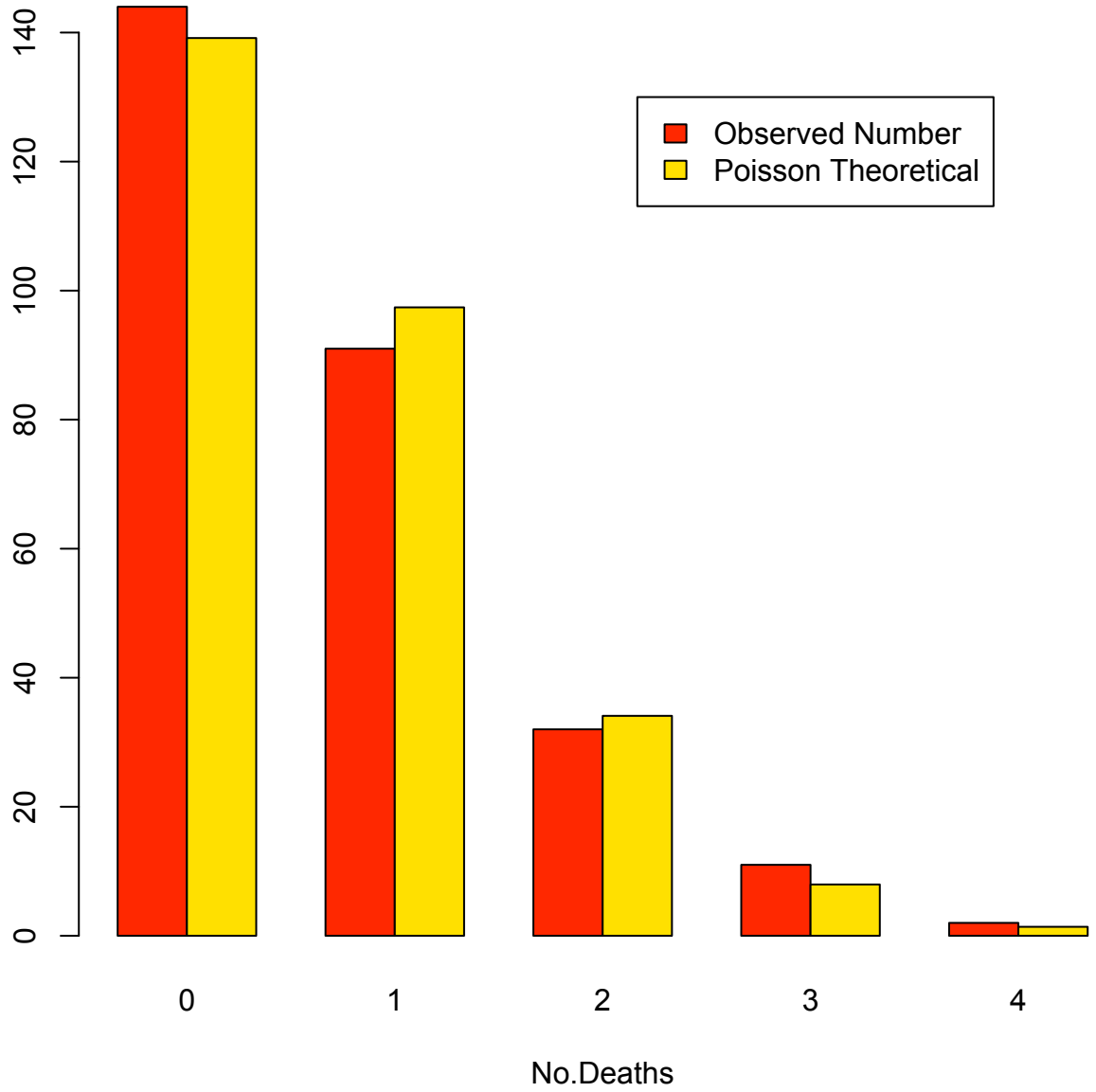
[Workspace restored from /Users/andrejstreibergs/.RData]

```
> ##### ENTER HORSE KICK DATA #####
> kick <- scan()
1: 144 91 32 11 2
6:
Read 5 items
> no <- 0:4
> N <- sum(kick)
> N
[1] 280
> ekick <- sum(no*kick)/N;ekick
[1] 0.7

> ##### PLOT HORSE KICK VS. THEORETICAL #####
> # Compute expected horse kicks. make sure sum is total kicks N.
> expfr <- dpois(0:4,ekick)*N
> sexpfr <- sum(expfr); sexpfr
[1] 279.7801

> clr <- rainbow(7)[c(1,2)]
> barplot(rbind(kick,expfr*N/sexpfr), names=0:4, beside=T, col=clr,
+ main = "Deaths from Horse Kicks", xlab = "No.Deaths")
> legend(8.5,130,legend=c("Observed Number","Poisson Theoretical"),fill=clr,bg="white")
```

Deaths from Horse Kicks



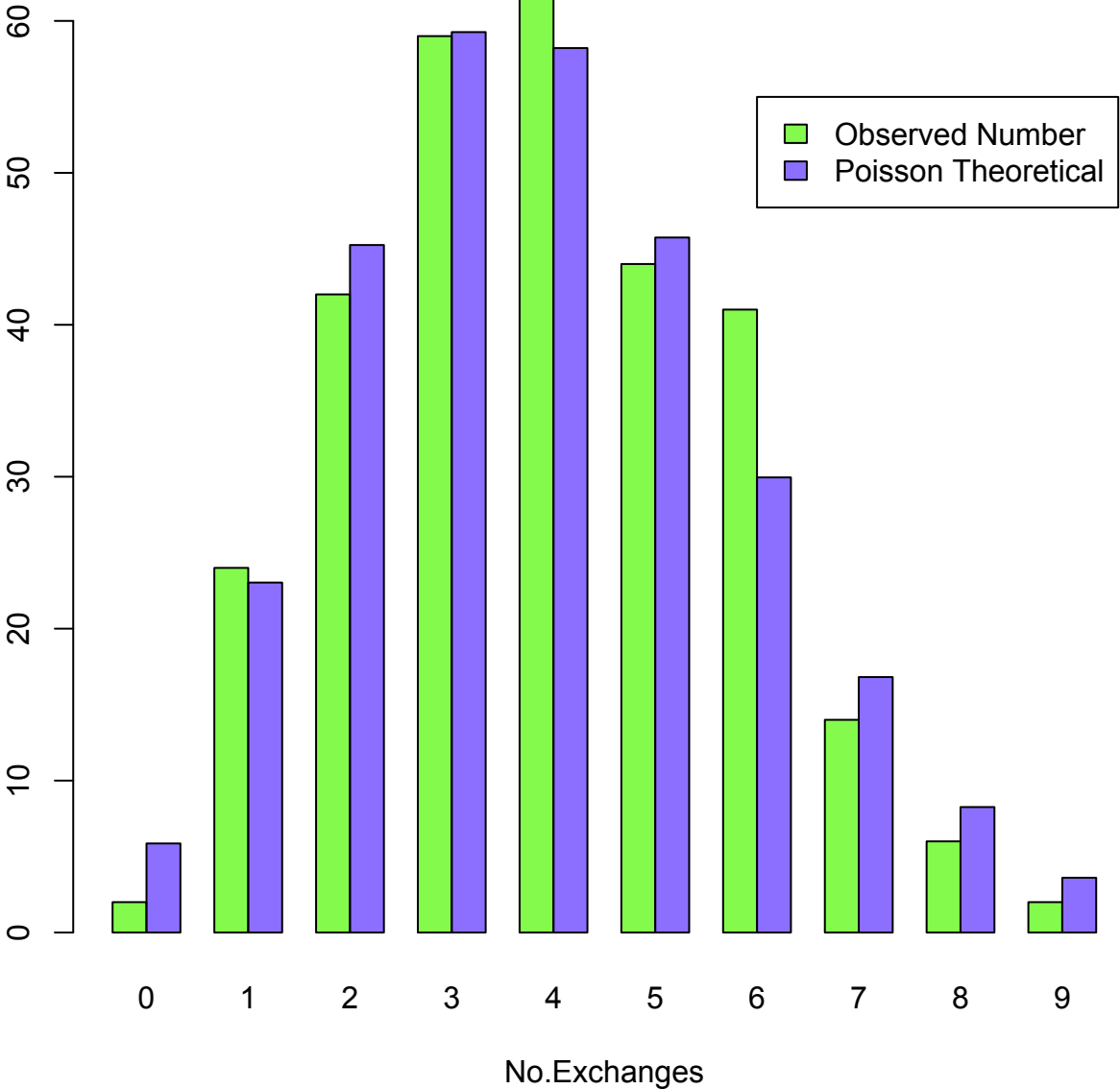
```

> ##### FIND 2-SIDED 95% CI FOR LAMBDA #####
> alpha <- .05
> za2 <- qnorm(alpha/2,lower.tail=F);za2
[1] 1.959964
>
> sr <- za2*sqrt(ekick/N+za2^2/(4*N^2))
> mc <- ekick + .5*za2^2/N
> c(mc-sr,mc+sr)
[1] 0.6086218 0.8050977
> ekick
[1] 0.7
> #M3074Horse1.pdf

> ##### ENTER CHROMATID EXCHANGE DATA #####
> # Devore 14.16
> xch <- scan()
1: 2 24 42 59 62 44 41 14 6 2
11:
Read 10 items
> xn <- 0:9
> xN = sum(xch); xN
[1] 296
> # Compute expected exchanges. Make sure it sums to total xN.
> exch <- sum(xn*xch)/xN; exch
[1] 3.929054
> ppois(9,exch)
[1] 0.9927656
> xP <- ppois(9,exch); xP
[1] 0.9927656
> xpred <- dpois(0:9,exch)*xN/xP
>
> clr <- rainbow(7,alpha=.7)[c(3,6)]
> barplot(rbind(xch,xpred), names = 0:9, beside=T, col=clr,
+ main = "Sister-Chromatid Exchanges", xlab="No.Exchanges")
> legend(20,55, legend = c("Observed Number","Poisson Theoretical"),
+ fill = clr, bg="white")
> # M3074Horse2.pdf

```

Sister-Chromatid Exchanges



```

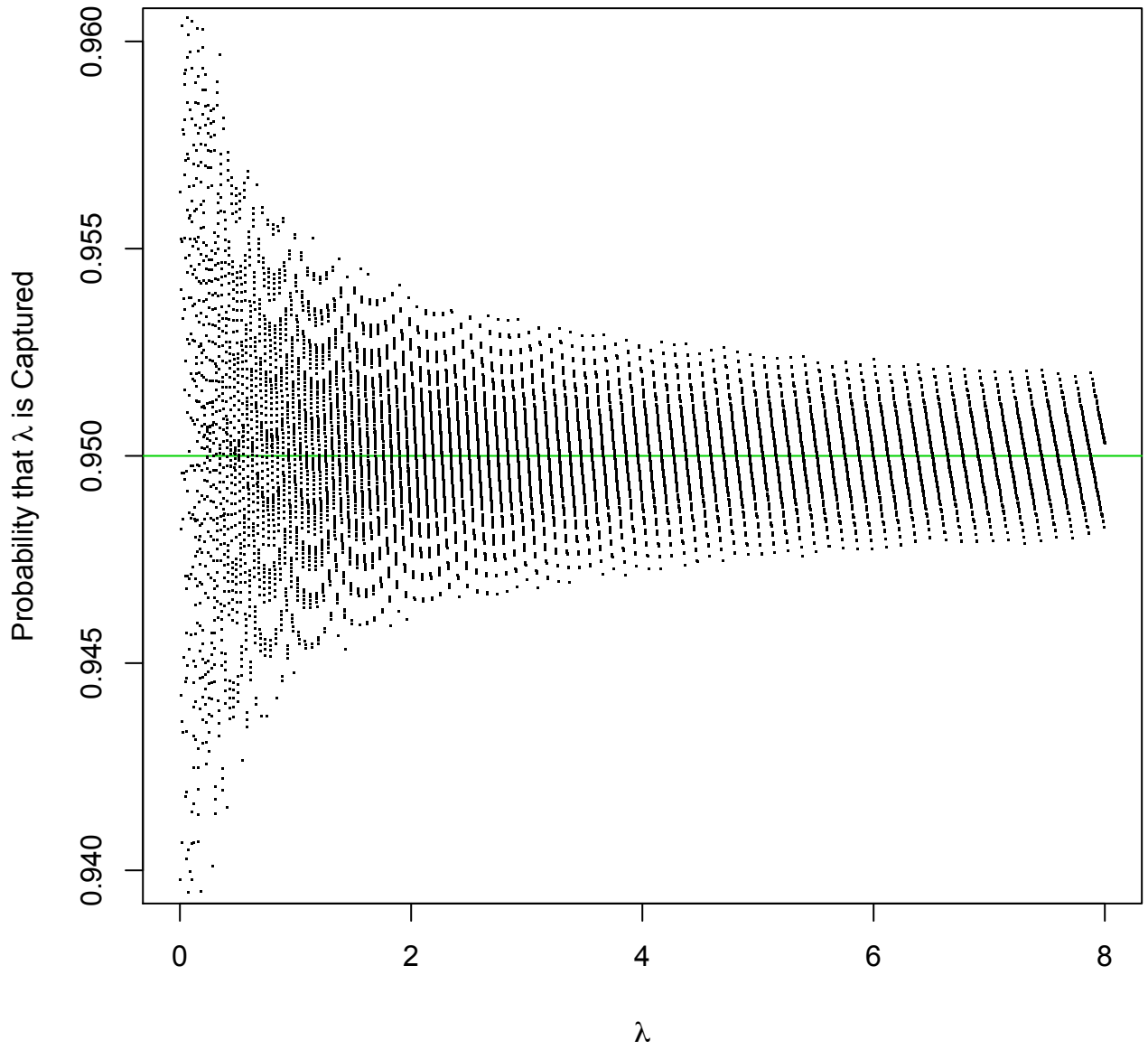
> ##### FIND 2-SIDED CI FOR LAMBDA #####
> alpha <- .01
> za2 <- qnorm(alpha/2, lower.tail=F); za2
[1] 2.575829
> sr <- za2*sqrt(exch/xN+za2^2/(4*xN^2))
> mc <- excl + .5*za2^2/xN
> c(mc-sr,mc+sr)
[1] 3.643283 4.237240
> excl
[1] 3.929054

> ##### COVERAGE PROBABILITY OF LAMBDA CI #####
> n <- 100
> alpha <- .05
> za2 <- qnorm(alpha/2,lower.tail=F);za2
[1] 1.959964
> cvr <- function(la){ppois(n*la +za2*sqrt(n*la),n*la)-
+ ppois(n*la -za2*sqrt(n*la),n*la)}

> las <- seq(0,8,1/1519)
> plot(las, cvr(las), type="n", ylim=c(.94,.96),
+ main = expression("Coverage Probabilities for CI on Poisson Parameter"),
+ xlab = expression(lambda), ylab = expression(paste("Probability that ",
+ lambda, " is Captured")))
> abline(h=1-alpha, col=3)
> points(las,cvr(las),ylim=c(.94,.96),pch=".")
> # M3074Horse3.pdf
>
> n <- 10
> cvr <- function(la){ppois(n*la +za2*sqrt(n*la),n*la)-
+ ppois(n*la -za2*sqrt(n*la),n*la)}
> curve(cvr,0,10)
> plot(las, cvr(las), type="n", ylim=c(.93,.97),
+ main = "Coverage Probabilities for CI on Poisson Parameter Samp.size=10",
+ xlab = expression(lambda), ylab = expression(paste("Probability that ",
+ lambda, " is Captured")))
> abline(h=1-alpha, col=3)
> points(las, cvr(las), ylim=c(.94,.96), pch=".")
> #M3074Horse4.pdf

```

Coverage Probabilities for CI on Poisson Parameter



Coverage Probabilities for CI on Poisson Parameter, Samp.size=10

