Importance sampling

Qinqing Li

2024-08-31

Three alternatives for Poisson parameter confidence intervals

Asymptotic Normality of the Maximum Likelihood Estimator (MLE)

Thm. : under mild assumptions, the difference between the MLE of θ and the true value of θ becomes approx. Gaussian matrix. With covariance matrix: inverse Fisher Information matrix.

$\hat{\theta}_{ML} - \theta \approx N(0, \tilde{H}(\theta)^{-1})$

Note information is inversely proportional to uncertainty (variance-covariance), the more information the less uncertainty. Therefore, we use the inverse Fisher matrix to quantify the uncertainty, instead of using the Fisher Information matrix to measure the amount of information.

when we have 1 param, converge to Normal distribution:

$\frac{\hat{\theta}_{ML} - \theta}{\sqrt{var(\hat{\theta}_{ML})}} \rightarrow N(0,1)$ and

$var(\hat{\theta}_{ML}) = [\tilde{H}(\theta)^{-1}]_{11}$

Interval construction

```
# Two-tailed test: 2.5% in each tail for 95% confidence level
# z1 <- qnorm(0.975)
# z2 <- qnorm(0.025)

CI1 <- function(y, alpha = 0.05) {
  lambda_hat <- sum(y)/length(y)
  lwr <- lambda_hat - sqrt(lambda_hat/length(y)) * qnorm(1 - alpha / 2)
  upr <- lambda_hat - sqrt(lambda_hat/length(y)) * qnorm(alpha / 2)
  return(c(lwr = lwr, upr = upr))
}

CI2 <- function(y, alpha = 0.05) {
  y_bar <- sum(y)/length(y)
  lwr <- (max(0,sqrt(y_bar)-qnorm(0.975)/(2*sqrt(length(y)))))**2
  upr <- (sqrt(y_bar) - qnorm(0.025)/(2*sqrt(length(y))))**2
  return(c(lwr = lwr, upr = upr))
}

CI3 <- function(y, alpha = 0.05) {
  lambda_hat <- sum(y)/length(y)
  n<-length(y)
  theta_hat<-log(mean(y))
```

```
  lwr <- log(mean(y)) - qnorm(1 - alpha / 2)/sqrt(n*exp(theta_hat))
  upr <- log(mean(y)) - qnorm(alpha / 2)/sqrt(n*exp(theta_hat))
  return(c(lwr = lwr, upr = upr))
}
```

```
set.seed(123)
y <- rpois(n = 10000, lambda = 7.5)
```

```
CI <- rbind(
  "Method 1" = CI1(y),
  "Method 2" = CI2(y),
  "Method 3" = CI3(y)
)
colnames(CI) <- c("Lower", "Upper")
```

Will all three methods always produce a valid interval? Ans: Not always, use n = 5 and lambda = 0.1, method 1 gives CI (-0.192, 0.592).

Bayesian credible intervals

side note: frequentist approach to statistical inference: there exists a true value for parameter θ, take repeated samplings to interpret the true value. Bayesian approach: θ is a r.v. Investigator has prior beliefs about θ before any observation of data, summarised in prior distribution $\pi(\theta)$. When data $Y = y$ is observed, the extra information is combined prior to obtaining posterior distribution $\pi(\theta|x)$ for θ given $Y = y$.

transforming between 2 distributions: $p(\theta) = p(\lambda)\frac{d\lambda(\theta)}{d\theta}$

Importance sampling

```
sim_sample <- function(y,m,a) {
  set.seed(123)
  n<- length(y)
  var_est <- 1/(1+n*mean(y))
  samples <- rnorm(m, mean = log(1 + sum(y)) - log(a + n), sd = sqrt(var_est))
  return(samples)
}
```

```
##  [1] 2.009159 2.010368 2.016912 2.011468 2.011683 2.017484 2.012896 2.006582
##  [9] 2.008697 2.009579
```

```
a <- 0.2
m <- 10
set.seed(12)
n<-length(y)
x <- rnorm(m, mean = log(1+sum(y)) - log(a + n), sd = 1 / sqrt(1 + sum(y)))
mean <-log(1+sum(y)) - log(a + n)
sd <- 1 / sqrt(1 + sum(y))
```

calculate unnormalised importance weights:

```
unnormalised_imp_w <- function(y,a = 1/5) {
  n <- length(y)
  #note log_posteror_distri is proportional to likelihood times prior,
  #where the constant of proportionality is chosen to make the total mass of the
```

```r
  #posterior distribution equal to 1
  log_posteror_distri <- (x * (1 + sum(y)) - (a + n) * exp(x))
  w <- log_posteror_distri - dnorm(m, mean = log(1+sum(y)) - log(a + n),
                                   sd = 1 / sqrt(1 + sum(y)), log = TRUE)
  return(w)
}
```

```r
exp(log(unnormalised_imp_w(y))-max(log(unnormalised_imp_w(y))))
```

```
##  [1] 0.9999996 0.9999995 0.9999998 0.9999998 0.9999992 1.0000000 1.0000000
##  [8] 0.9999999 1.0000000 1.0000000
```

```r
log_weights <- (x * (1 + sum(y)) - (a + n) * exp(x)) -
 dnorm(x, mean = log(1 + sum(y)) - log(a + n), sd = 1 / sqrt(1 + sum(y)), log = TRUE)
weights <- exp(log_weights - max(log_weights))
```

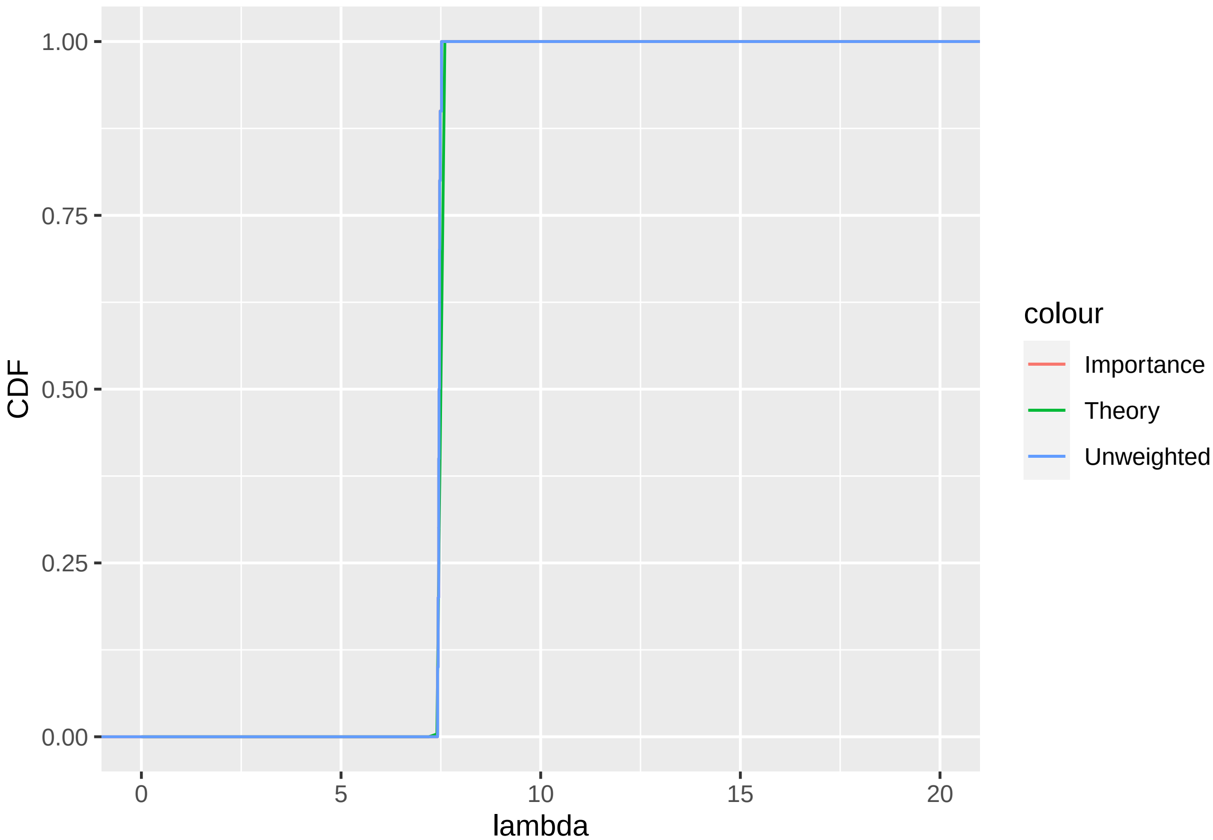
```r
theta_interval <- wquantile(x, probs = c(0.025, 0.975), weights = weights)
theta_interval
```

```
## [1] 2.004326 2.016032
```

```r
lambda_interval <- exp(theta_interval)
lambda_interval
```

```
## [1] 7.421092 7.508470
```

```r
ggplot(data.frame(lambda = exp(x), weights = weights)) +
  xlim(0, 20) + ylab("CDF") +
  geom_function(fun = pgamma, args = list(shape = 1 + sum(y), rate = a + n),
                mapping = aes(col = "Theory")) +
  stat_ewcdf(aes(lambda, weights = weights, col = "Importance")) +
  stat_ecdf(aes(lambda, col = "Unweighted"))
```



1.00
0.75
0.50
0.25
0.00
CDF
0
5
10
15
20
lambda
colour
Importance
Theory
Unweighted