Estudo sobre modelos log-lineares Poisson: Verificação de estimativa pontual e intervalar (tipo link) para as médias

André Savassi
Kenzo Bontempo
Márcio Antônio

UFMG

01/10/2025

Sumário

Introdução

- **Modelos log-lineares Poisson:** principais ferramentas para modelar variáveis de contagem.
  - Aplicações em epidemiologia, ciências sociais e controle de qualidade.
- A média é expressa como **função exponencial das covariáveis**, garantindo positividade das estimativas.
- **Objetivo do estudo:**
  - Verificar **estimativas pontuais e intervalares** da média do modelo Poisson.
  - Utilizar **simulações de Monte Carlo** para avaliar desempenho dos estimadores.

Introdução

- **Métricas analisadas:**
  - Viés relativo
  - Erro padrão
  - Taxa de cobertura dos intervalos de confiança
- **Comparações adicionais:**
  - Distribuições **Binomial Negativa** e **Bell**, avaliando robustez e sobredispersão.

Gerando amostras de distribuições - Poisson

Primeiramente, iremos amostrar da Distribuição Poisson, que possui a função densidade descrita por:

$$f(y;\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \;\; y = 0, 1, \ldots, n$$

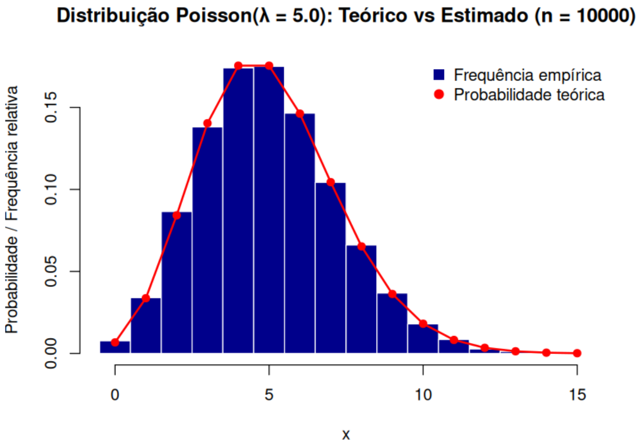
Referencia-se que: $\mathbb{E}(Y) = Var(Y) = \lambda$

Amostrador Poisson via Aceitação e Rejeição

```
^^Ipoisson_ar_1 <- function(lambda) {
^^I^^If_pois <- function(k, lambda) {
^^I^^I^^Ireturn(exp(k * log(lambda) - lambda - lgamma(k
    + 1)))
^^I^^I}
^^I^^Ifmax <- f_pois(floor(lambda), lambda)
^^I^^Ikmax <- qpois(0.9999, lambda = 5)
^^I^^Irepeat {
^^I^^I^^Iy <- sample(0:kmax, 1)
^^I^^I^^Ify <- f_pois(y, lambda)
^^I^^I^^Igy <- 1 / (kmax + 1)
^^I^^I^^Iu <- runif(1)
^^I^^I^^Iif (u < fy / (fmax)) {
^^I^^I^^I^^Ireturn(y)
^^I^^I^^I}}}
^^Ipoisson_ar <- function(lambda, n = 1) {
^^I^^Ireplicate(n, poisson_ar_1(lambda))
^^I}^^I^^I
```

Código.R

Gráficos - Poisson AR



Gerando amostras de distribuições - Binomial Negativa

Segue que a densidade da distribuição Binomial Negativa é descrita por:

$$f(y; r, p) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$$

onde r é o total de sucessos e p é a probabilidade desses sucessos. Note que $Y \sim BN(r, p)$ possui esperança e variância a seguir:

$$\mathbb{E}(Y) = \frac{r}{p} \; ; \; Var(Y) = \frac{r(1-p)}{p^2}$$

Amostrador BN via Amostragem por Importância

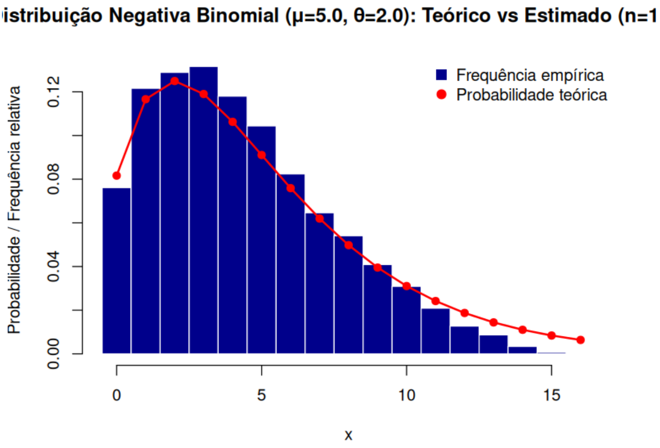
```
rnegbinom <- function(m, n, mu, theta) {
^^If_negbinom <- function(x, mu, theta) {
^^I^^Ilog_f <- lgamma(x + theta) -
^^I^^Ilgamma(theta) -
^^I^^Ilgamma(x + 1) +
^^I^^Itheta * log(theta / (mu + theta)) +
^^I^^Ix * log(mu / (mu + theta))
^^I^^Ireturn(exp(log_f))
^^I}
  g_pois <- function(k, lambda) {
^^Ireturn(exp(k * log(lambda) - lambda - lgamma(k + 1))
    )
}

^^I^^I^^I^^I
```

▸ Código.R

Amostrador BN via Amostragem por Importância

```
  source("rpoisson.R")

y <- rpois(n = m, lambda = 0.9 * mu)

pesos <- f_negbinom(y, mu = mu, theta = theta) / g_pois(y, lambda
      = 0.9 * mu)
pesos <- pesos / sum(pesos)

amostra <- sample(y, n, F, pesos)

return(amostra)
}
^^I^^I^^I
```

Gráficos - BN Amostragem por Importância



Gerando amostras de distribuições - Bell

Segue que a densidade da distribuição Bell é descrita por:

$$f(y;\theta) = \frac{\theta e^{e^{\theta}+1}}{y!} B_Y, \;\; y = 0, 1, \ldots, n; \;\; \theta > 0$$

Onde o termo B_Y corresponde ao número de Bell, dados por:

$$B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}, \text{iniciando com } B_0 = B_1 = 1$$

A média e variância de $Y \sim Bell(\theta)$ é:

$$\mathbb{E}(Y) = \theta e^{\theta} \; ; \;\; Var(Y) = \theta(1+\theta)e^{\theta}$$

Amostrador Bell via Transformação Inversa

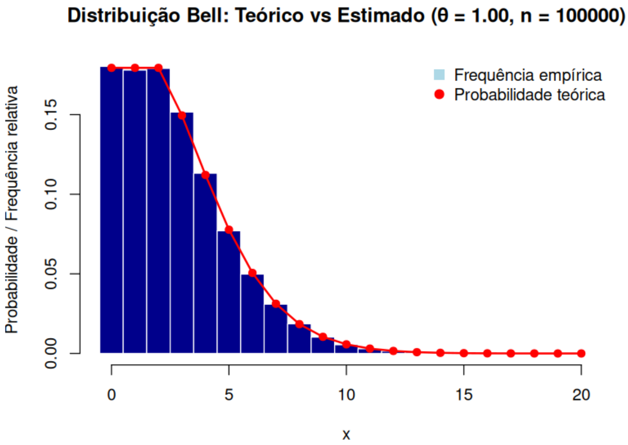
```
^^I^^I# Importado de bellreg
^^I^^Idbell <- function(x, theta, log = FALSE) {
^^I^^I^^IBx <- c()
^^I^^I^^Ifor (i in 1:length(x)) {
^^I^^I^^I^^IBx[i] <- numbers::bell(x[i])
^^I^^I^^I}
^^I^^I^^Ilf <- x * log(theta) - exp(theta) + 1 + log(Bx
    ) - lgamma(x + 1)
^^I^^I^^Iif (log == TRUE) {
^^I^^I^^I^^Ireturn(lf)
^^I^^I^^I} else {
^^I^^I^^I^^Ireturn(exp(lf))
^^I^^I^^I}
^^I^^I}^^I^^I
^^I^^I^^I^^I
```

Código.R

Amostrador Bell via Transformação Inversa

```
^^I^^Irbell_aux <- function(theta) {
^^I^^I^^Isapply(theta, function(t) {
^^I^^I^^I^^Iu <- runif(1, 0, 1)
^^I^^I^^I^^Ii <- 0
^^I^^I^^I^^Ipr <- dbell(0, t); Fx <- pr
^^I^^I^^I^^Iwhile (u >= Fx) {
^^I^^I^^I^^I^^Ii <- i + 1
^^I^^I^^I^^I^^Ipr <- dbell(i, t)
^^I^^I^^I^^I^^IFx <- Fx + pr}
^^I^^I^^I^^Ireturn(i)
^^I^^I^^I}))}
^^I^^Irbell <- function(n, theta) {
^^I^^I^^Iif (length(theta) == 1) {
^^I^^I^^I^^Ireturn(replicate(n, expr = rbell_aux(theta), simplify
    = TRUE))
^^I^^I^^I} else {
^^I^^I^^I^^Ireturn(rbell_aux(theta))
^^I^^I^^I}}
```

Gráficos - Bell Inversa



Implementação do Algoritmo

- 1. Gere covariáveis contínuas e discretas e as organize na matriz $X_{(n+m)xp}$;
- 2. Escolha arbitrariamente um vetor β;
- 3. Calcule $\mu_i = exp(\eta_i)$, onde $\eta_i = X_i^T\beta$, $\forall i = 1, \ldots, n+m$;
- 4. Gere Y_i de Poisson(μ_i), $\forall i = 1, \ldots, n$;
- 5. Avaliar o modelo usando as n primeiras amostras;
- 6 .Calcular estimação pontual e intervalar da média mu_i para as m últimas amostras;
- 7. Compare os resultados com os μ_i gerados no item(3), a partir do víes relativo e CP cobertura;
- 8. Repita de 1 até 7 até alcançar o número de repetições desejadas.

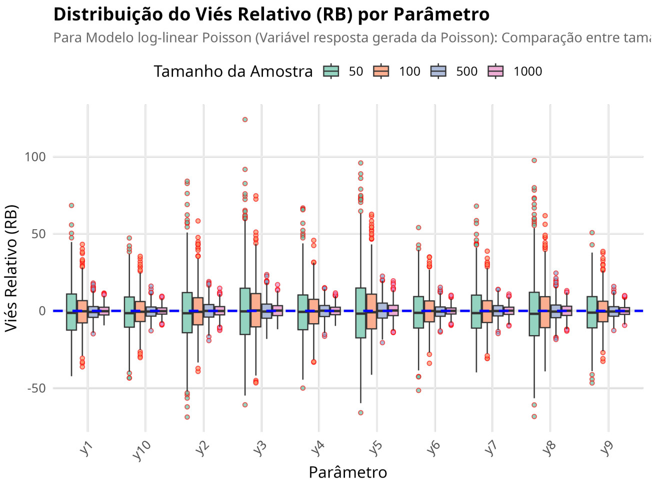
Monte Carlo - Poisson

Código.R

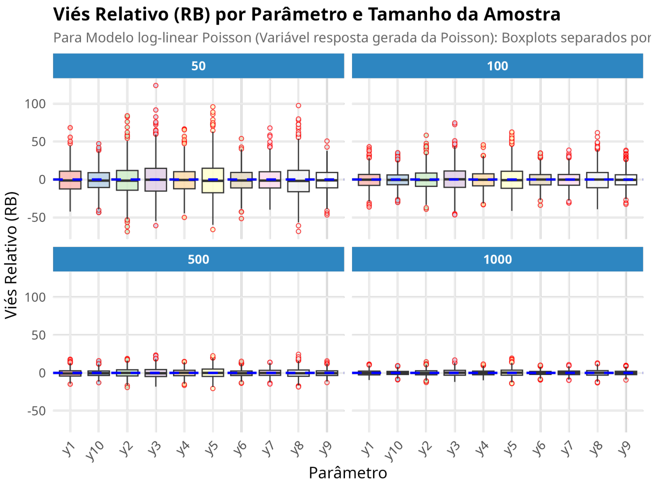
Tabelas - MC Poisson

Tamanho da Amostra	Parâmetro	Valor Real	Estimativa Média	Erro Padrão	Desvio Padrão	Viés Relativo	Limite Inferior	Limite Superior	Cobertura
100	y1	3.362	3.354	0.014	0.386	-0.225	2.673	4.211	0.950
100	y10	3.798	3.790	0.010	0.378	-0.210	3.119	4.607	0.958
100	y2	3.491	3.492	0.020	0.481	0.029	2.664	4.581	0.952
100	y3	5.447	5.497	0.025	0.914	0.934	4.036	7.494	0.937
100	y4	2.513	2.507	0.014	0.285	-0.259	1.998	3.146	0.952
100	y5	3.980	3.997	0.029	0.682	0.424	2.873	5.565	0.949
100	y6	4.220	4.217	0.011	0.440	-0.057	3.447	5.162	0.950
100	y7	3.097	3.093	0.012	0.334	-0.151	2.501	3.825	0.956
100	y8	4.512	4.514	0.021	0.658	0.039	3.401	5.998	0.959
100	y9	3.845	3.840	0.011	0.392	-0.148	3.143	4.692	0.958

Gráficos - MC Poisson



Gráficos - MC Poisson



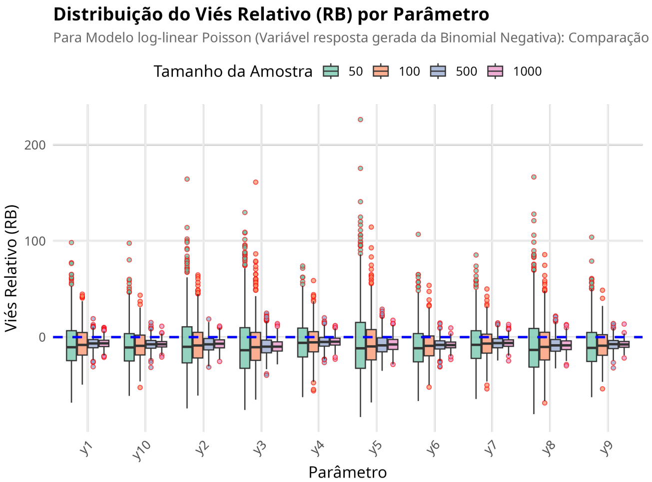
Monte Carlo - Binomial Negativa

Código.R

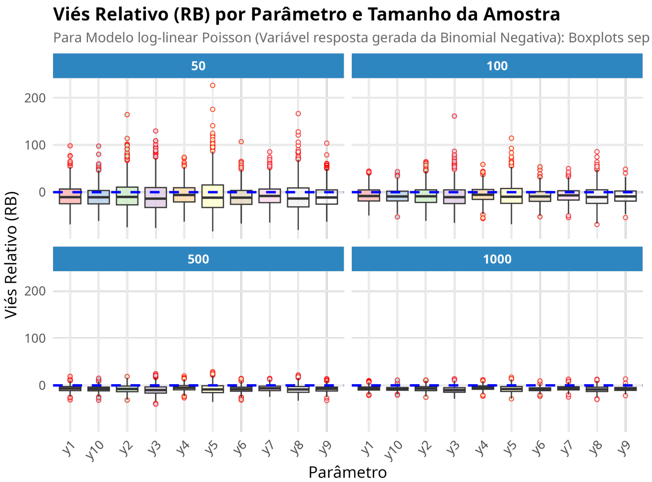
Tabelas - MC Binomial Negativa

Tamanho da Amostra	Parâmetro	Valor Real	Estimativa Média	Erro Padrão	Desvio Padrão	Viés Relativo	Limite Inferior	Limite Superior	Cobertura
100	y1	3.362	3.142	0.015	0.562	-6.546	2.485	3.975	0.783
100	y10	3.798	3.494	0.011	0.560	-8.012	2.853	4.280	0.734
100	y2	3.491	3.227	0.022	0.711	-7.559	2.438	4.277	0.775
100	y3	5.447	4.978	0.028	1.286	-8.607	3.612	6.867	0.782
100	y4	2.513	2.397	0.015	0.399	-4.626	1.897	3.029	0.820
100	y5	3.980	3.715	0.032	0.994	-6.660	2.642	5.231	0.783
100	y6	4.220	3.853	0.012	0.654	-8.686	3.123	4.757	0.718
100	y7	3.097	2.902	0.013	0.457	-6.317	2.331	3.614	0.796
100	y8	4.512	4.120	0.024	0.966	-8.696	3.071	5.535	0.760
100	y9	3.845	3.532	0.012	0.581	-8.142	2.869	4.351	0.731

Gráficos - MC Binomial Negativa



Gráficos - MC Binomial Negativa



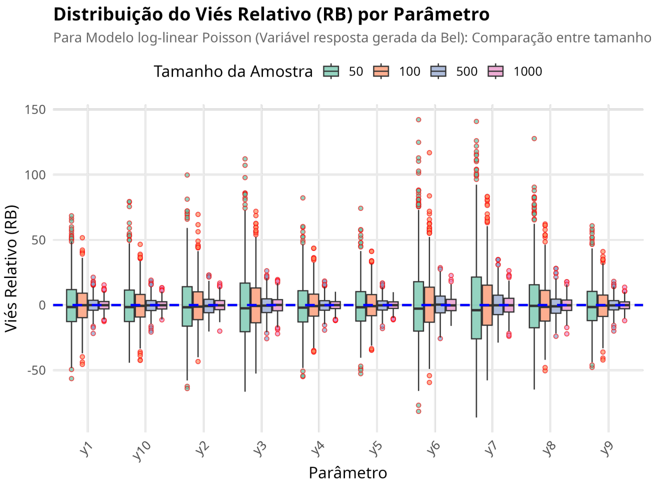
Monte Carlo - Bell

Código.R

Tabelas - MC Bell

Tamanho da Amostra	Parâmetro	Valor Real	Estimativa Média	Erro Padrão	Desvio Padrão	Viés Relativo	Limite Inferior	Limite Superior	Cobertura
100	y1	3.362	3.374	0.014	0.578	0.367	2.691	4.233	0.820
100	y10	3.798	3.785	0.010	0.568	-0.326	3.116	4.601	0.812
100	y2	3.491	3.482	0.020	0.715	-0.243	2.658	4.568	0.823
100	y3	5.447	5.531	0.026	1.370	1.547	4.064	7.534	0.796
100	y4	2.513	2.530	0.014	0.422	0.676	2.018	3.174	0.824
100	y5	3.980	4.032	0.029	0.972	1.321	2.904	5.607	0.823
100	y6	4.220	4.206	0.011	0.666	-0.319	3.438	5.149	0.816
100	y7	3.097	3.113	0.012	0.506	0.506	2.519	3.848	0.815
100	y8	4.512	4.511	0.022	0.964	-0.020	3.403	5.989	0.822
100	y9	3.845	3.830	0.011	0.586	-0.388	3.136	4.680	0.816

Gráficos - MC Bell



Gráficos - MC Bell