Homework 1

João Carabetta, `Estatística`

Exercício 1

(a) Assuma que $X_1 \sim \text{Poisson}(\mu)$ e encontre a transformação estabilizadora de variância para a média.

Dado que a $X_1 \sim \text{Poisson}(\mu)$ então $Var(\mu) = \sigma^2(\mu) = \mu$ logo

$$\sigma(\mu) = \pm\sqrt{\mu}.. \tag{1}$$

Portanto,

$$|g'(\mu)|\sigma(\mu) = c \Rightarrow |g'(\mu)|\sqrt{\mu} = c \Rightarrow g'(\mu) = \pm|\frac{c}{\sqrt{\mu}}| \tag{2}$$

Assim,

$$g(\mu) = \pm 2c\sqrt{\mu} + d, d \in R \tag{3}$$

(b) Mostre que $\sqrt{n}(g(Y_n) - g(\mu)) \to N(0, 1/4)$.

Sabendo que $\sqrt{n}(g(Y_n) - g(\mu)) \to N(0, \sigma(\mu)^2 g'(\mu)^2)$, então, devemos escolher c tal que, $c = \sigma(\mu)g'(\mu) = 1/2$

(c) Qual a importância prática desta transformação ? (explique sucintamente ou apresente um exemplo)
Essa transformação permite estimar uma estatística de uma certa distribuição. Por exemplo, é útil para estimar a proporção de entre membros de uma população com características A, B, C.

Exercício 2

(a) Mostre que $\sqrt{n}(g(Y_n) - g(\mu)) = o_p(1)$.
Supondo que existe $g(y)$ então, podemos expandir g em Taylor torno de Y_n :

$$g(Y_n) = g(\mu) + g'(\mu)(Y_n - \mu) + \frac{g''(\mu)}{2}(Y_n - \mu)^2 + R, \tag{4}$$

onde R é o resíduo da expansão. Assim, se operando os termos e multiplicando por $\sqrt{n}$ temos :

$$\sqrt{n}(g(Y_n) - g(\mu)) = \frac{g''(\mu)}{2}\frac{\sqrt{n}}{n^2}\Big(\sum_{i=0}^{\infty} X_i - n\mu\Big)^2 + \sqrt{n}R. \tag{5}$$

Se observarmos os termos, veremos que $g''(\mu) = O(1)$, $\frac{\sqrt{n}}{n^2} = o(1)$, $\Big(\sum_{i=0}^{\infty} X_i - n\mu\Big)^2 = o_P(1)$ e $\sqrt{n}R = o_p(1)$. Portanto, o produto deles é $o_p = 1$. Finalmente,

$$\sqrt{n}(g(Y_n) - g(\mu)) = o_p(1) \tag{6}$$

(b) Mostre que $n(g(Y_n) - g(\mu)) \xrightarrow{d} \frac{1}{2}g''(\mu)\sigma^2 V$, onde $V \sim \chi^2_1$.
Continuando da Eq.4, mas dessa vez multiplicando por n os dois lados e por $\frac{\sigma^2}{\sigma^2}$ o lado direito temos,

$$n(g(Y_n) - g(\mu)) = \sigma^2 \frac{g''(\mu)}{2} \frac{n\big(Y_n - \mu\big)^2}{\sigma^2} + nR. \tag{7}$$

Sabendo que $\frac{n\left(Y_n-\mu\right)^2}{\sigma^2} \xrightarrow{d} \chi^2_1$ pelo Teorema Central do Limite e $nR = o_p(1)$, então, pelo Teorema de Slutsky temos,

$$n(g(Y_n) - g(\mu)) \xrightarrow{d} \frac{1}{2}g''(\mu)\sigma^2\chi^2_1 \tag{8}$$

(c) Suponha que $X_1 \sim Bernoulli(p)$. Ache a distribuição de $g(Y_n) = Y_n(1 - Y_n)$.
Temos que $EX_1 = p$ e $Var(X_1) = p(1 - p)$ para uma distribuição Bernoulli. Assim, podemos usar $g(Y_n) = Y_n(1 - Y_n)$ para estimar a $Var(X_1) = p(1 - p)$. Assim,

$$g'(Y_n) = 1 - 2Y_n \tag{9a}$$
$$g''(Y_n) = -2 \tag{9b}$$

mas, para como Y_n estima p, $g'(p) = 0$ quando $p = 1/2$. Portanto, para $p \neq 1/2$ usamos Método Delta de Primeira Ordem,

$$\sqrt{n}(g(Y_n) - g(p)) \xrightarrow{d} N(0, \sigma^2 g'(p)^2) \Rightarrow \tag{10a}$$
$$g(Y_n) \xrightarrow{d} N(g(p), \frac{\sigma^2 g'(p)^2}{n}) \Rightarrow \tag{10b}$$
$$g(Y_n) \xrightarrow{p} p(1 - p) \tag{10c}$$

para $p = 1/2$ é melhor usar o Método Delta de Segunda Ordem para aproximar a distribuição. Assim,

$$n(g(Y_n) - g(p)) = \frac{1}{2}g''(\mu)\sigma^2 \frac{n\big(Y_n - \mu\big)^2}{\sigma^2} \Rightarrow \tag{11a}$$
$$ng(Y_n) - n[p(1 - p)] = \frac{1}{2}(-2)p(1 - p)\frac{n\big(Y_n - \mu\big)^2}{\sigma^2} \Rightarrow \tag{11b}$$
$$g(Y_n) = p(1 - p)\Big(1 - \frac{1}{n}\frac{n\big(Y_n - \mu\big)^2}{\sigma^2}\Big) \xRightarrow{Slutsky} \tag{11c}$$
$$g(Y_n) \xrightarrow{d} p(1 - p)\Big(1 - \frac{1}{n}\chi^2_1\Big) \Rightarrow \tag{11d}$$
$$g(Y_n) \xrightarrow{p} p(1 - p) \tag{11e}$$

Exercício 3

(a) $\hat{p} = \sum_{i=1}^{2} 4x_i = 11/24$.
Usando o método **Delta** :

Para $p \neq 1/2$, podemos usar a Eq.10b, com $\sigma^2 = p(1-p)$ e $g'(p) = (1-2p)$

$$g(P_n) \xrightarrow{d} N(g(p), \frac{\sigma^2 g'(p)^2}{n}) \Rightarrow \tag{12a}$$

$$g(P_n) \xrightarrow{d} N\left(p(1-p), \frac{p(1-p)(1-2p)^2}{n}\right) \Rightarrow \tag{12b}$$

$$g(P_n) \xrightarrow{d} N\left(0.2482, 0.0003\right) \tag{12c}$$

Usando o método **Bootstrap** :

Para o método Bootstrap, confira no GitHub. Os resultados são $Eg(P_n) = 0.23$ e $Var(g(P_n)) = 0.01$.

(b) $\hat{p} = \sum_{i=1}^{2} x_i = 12/24$.

Usando o método **Delta** :

Agora, para $p = 1/2$ temos que usar Eq.11d. Assim, a esperança é,

$$Eg(Y_n) = E\left(p(1-p)\left(1 - \frac{\chi_1^2}{n}\right)\right) \Rightarrow \tag{13a}$$

$$Eg(Y_n) = p(1-p)(1-1/n) \Rightarrow \tag{13b}$$

$$Eg(Y_n) = 0.239583 \tag{13c}$$

e a variância é,

$$Var(g(Y_n)) = Var\left(p(1-p)\left(1 - \frac{\chi_1^2}{n}\right)\right) \Rightarrow \tag{14a}$$

$$Var(g(Y_n)) = p^2(1-p)^2 Var(1 - \frac{\chi_1^2}{n}) \Rightarrow \tag{14b}$$

$$Var(g(Y_n)) = \frac{2p^2(1-p)^2}{n^2} \Rightarrow \tag{14c}$$

$$Var(g(Y_n)) = 0.001 \tag{14d}$$

Logo, $g(Y_n) = 0.240 \pm 0.001$.

Usando o método **Bootstrap** :

Para o método Bootstrap, confira no GitHub. Os resultados são $Eg(P_n) = 0.23$ e $Var(g(P_n)) = 0.01$.

(c) Suponha que o valor verdadeiro de $p = 0.4$. Calcule a variância "exata" usando simulação e discuta os resultados acima. Qual método você se sentiria mais confiante em empregar ?

Se a $p = 0.4$, então a variância 'exata' pode ser calculada por $Var(p) = p(1-p) = 0.24$. Ambos os métodos me retornaram valores próximos a 0.24 nas duas situações. A única grande diferença foi na variância pequena que o método Delta retorna, o que me faz ou i) desconfiar da minha metodologia ou ii) acreditar que ele seja excessivamente preciso. Isso quer dizer que o método Delta estima erros em ordens de grandeza menores que a da casa decimal relevante, fazendo com que o método seja pouco útil para estimar o valor 'exato' com confiabilidade. Os valores do Bootstrap tinham variância na casa decimal de relevância, que é a segunda, e, por isso, pareceu ser um estimador mais estável.