Disciplina: MAE0212: Introdução a Probabilidade e Estatística II
Professora: Silvia Ferrari
Nome: **Nota**:

Use os espaços em branco, inclusive o verso das páginas, para responder às questões. Mostre os desenvolvimentos, não apenas as respostas.

Provinha 2 - *Gabarito*

1. Seja $(X_1, \ldots, X_n)$ uma amostra aleatória simples de uma variável aleatória X com distribuição de Poisson de parâmetro $\lambda > 0$. A função densidade de probabilidade de X é dada por

$$\mathbb{P}_\lambda(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots.$$

Lembre que $\mathbb{E}(X) = \lambda$ e Var$(X) = \lambda$.

a) (1 ponto) Obtenha o estimador de máxima verossimilhança de λ e verifique se ele é não viciado.

b) (1.5 ponto) Suponha que n seja grande para possibilitar o uso do Teorema Limite Central. Obtenha um intervalo de confiança para λ com coeficiente de confiança (aproximadamente) γ. Se a média amostral de uma amostra observada de tamanho 100 é 12, qual é a estimativa de máxima verossimilhança de λ? Qual é o intervalo de confiança se $\gamma = 92\%$?

Solução:

Seja $X \sim Poisson(\lambda)$.

a) A função de verossimilhança de λ, dada a amostra observada $(x_1, \ldots, x_n)$, é dada por

$$\begin{aligned}
L(\lambda) &= \mathbb{P}_\lambda(X_1 = x_1, \ldots, X_n = x_n) \\
&= \mathbb{P}_\lambda(X_1 = x_1) \cdots \mathbb{P}_\lambda(X_n = x_n) \quad \text{(pois } X_1, \ldots, X_n \text{ são independentes)} \\
&= \prod_{i=1}^n \mathbb{P}_\lambda(X = x_i) \quad \text{(pois } X_1, \ldots, X_n \text{ têm a mesma distribuição de X)} \\
&= \prod_{i=1}^n \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \\
&= \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}.
\end{aligned}$$

Assim, o logaritmo da função de verossimilhança é dado por

$$\ell(\lambda) = -n\lambda + \log(\lambda)\sum_{i=1}^n x_i - \log\left(\prod_{i=1}^n x_i!\right).$$

Note que o objetivo é encontrar o valor $\widehat{\lambda}$ que maximiza $\ell(\lambda)$. Dessa forma, temos que

$$\frac{\mathrm{d}\ell(\lambda)}{\mathrm{d}\lambda} = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}.$$

A estimativa de máxima verossimilhança de λ, $\widehat{\lambda}$, é tal que $\mathrm{d}\ell(\lambda)/\mathrm{d}\lambda$ avaliada em $\lambda = \widehat{\lambda}$ é igual a zero. Assim, temos que

$$-n + \frac{\sum_{i=1}^{n} x_i}{\widehat{\lambda}} = 0 \quad \Rightarrow \quad \widehat{\lambda} = \bar{x}.$$

Além disso, $\mathrm{d}^2\ell(\lambda)/\mathrm{d}\lambda^2 = -\sum_{i=1}^{n} x_i/\lambda^2 < 0, \forall \lambda > 0$, em particular, para $\lambda = \widehat{\lambda}$. Assim, $\bar{x}$ maximiza a função $\ell(\lambda)$ e, portanto, o estimador de máxima verossimilhança de λ é $\bar{X}$. Adicionalmente, note que

$$\mathbb{E}_{\lambda}(\bar{X}) = \mathbb{E}_{\lambda}(X) = \lambda.$$

Logo, o estimador de máxima verossimilhança $\bar{X}$ é não viciado para λ.

b) Pelo Teorema Limite Central, quando n é suficientemente grande, temos que $\bar{X} \sim N\left(\lambda, \frac{\lambda}{n}\right)$, de modo que

$$\frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \sim N(0,1).$$

Dessa forma,

$$\mathbb{P}\left(-z_{\gamma} \leq \frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \leq z_{\gamma}\right) = \gamma,$$

em que z_{γ} é tal que $\mathbb{P}(-z_{\gamma} \leq Z \leq z_{\gamma}) = \gamma$, sendo $Z \sim N(0,1)$. Assim,

$$\begin{aligned} &\mathbb{P}\left(-z_{\gamma} \leq \frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \leq z_{\gamma}\right) = \gamma \\ \Rightarrow\ &\mathbb{P}\left(-z_{\gamma}\sqrt{\frac{\lambda}{n}} \leq \bar{X} - \lambda \leq z_{\gamma}\sqrt{\frac{\lambda}{n}}\right) = \gamma \\ \Rightarrow\ &\mathbb{P}\left(\bar{X} - z_{\gamma}\sqrt{\frac{\lambda}{n}} \leq \lambda \leq \bar{X} + z_{\gamma}\sqrt{\frac{\lambda}{n}}\right) = \gamma. \end{aligned}$$

Substituindo λ por $\widehat{\lambda} = \bar{X}$ nos extremos do intervalo, temos

$$\mathrm{IC}(\lambda;\gamma) = \left[\bar{X} - z_{\gamma}\sqrt{\frac{\bar{X}}{n}}, \bar{X} + z_{\gamma}\sqrt{\frac{\bar{X}}{n}}\right].$$

Considerando que de uma amostra de tamanho 100 foi observado $\bar{x} = 12$, então a estimativa de máxima verossimilhança de λ é 12. Ainda, o intervalo de confiança para λ considerando essa amostra observada e $\gamma = 92\%$ é dado por

$$\left[12 - 1.75 \cdot \sqrt{0.12}, 12 + 1.75 \cdot \sqrt{0.12}\right] = [11.39, 12.61].$$

2. De uma população normal com $\sigma = 5$, retira-se uma amostra aleatória simples de tamanho 50, obtendo-se $\overline{x} = 42$.

   a) (1 ponto) Obtenha o intervalo de confiança para a média (μ) com coeficiente de confiança de 93%.

   b) (1.5 ponto) Para que o comprimento do intervalo de 93% confiança fosse 1, qual deveria ser o tamanho da amostra?

   Solução:

   Seja $X \sim N(\mu, 5^2)$. Considere uma amostra aleatória simples de tamanho 50 e note que $\bar{X} \sim N(\mu, 5^2/50)$.

   a) O intervalo de confiança observado para μ com coeficiente de confiança γ é dado por

   $$\text{IC}(\mu;\gamma) = \left[\bar{x} - z_\gamma \frac{\sigma}{\sqrt{n}}, \bar{X} + z_\gamma \frac{\sigma}{\sqrt{n}}\right],$$

   em que z_γ é tal que $\mathbb{P}(-z_\gamma \leq Z \leq z_\gamma) = \gamma$, sendo $Z \sim N(0,1)$. Para $\gamma = 93\%$, temos que $z_\gamma = 1.81$, assim temos que o intervalo de confiança para μ com $\gamma = 93\%$ é dado por

   $$\left[42 - 1.81\frac{5}{\sqrt{50}}, 42 + 1.81\frac{5}{\sqrt{50}}\right] = [40.72, 43.28].$$

   b) Para que o comprimento do intervalo seja igual a 1, temos que

   $$42 + 1.81\frac{5}{\sqrt{n}} - 42 + 1.81\frac{5}{\sqrt{n}} = 1 \Rightarrow 2 \cdot 1.81\frac{5}{\sqrt{n}} = 1 \Rightarrow n = (2 \cdot 1.81 \cdot 5)^2 = 327.61.$$

   Assim, o tamanho amostral necessário para que o intervalo tenha comprimento 1 é $n = 328$.

3. Um biólogo está estudando o problema de árvores de uma grande cidade estarem contaminadas por uma praga e, inicialmente, pretende estimar a proporção p de árvores contaminadas. Para essa finalidade o biólogo vai selecionar uma amostra aleatória das árvores da cidade.

   a) (1 ponto) Qual deve ser o tamanho da amostra para que o erro da estimativa seja no máximo 0,06 com probabilidade 0,94?
   b) (0.5 ponto) Tendo a informação de que essa proporção p é no máximo 35%, é possível considerar em (a) uma amostra de tamanho menor? Se sim, de quanto? Se não, por quê?
   c) (1 ponto) Se em uma amostra de 180 árvores foi verificada a presença de praga em 36 delas, dê uma estimativa pontual para p e, com base nela, construa um intervalo de 95% de confiança para essa proporção populacional.

   Solução:

   Considere a variável aleatória X que assume o valor 1 se uma árvore selecionada ao acaso está contaminada e 0 caso contrário. Note que $X \sim Bernoulli(p)$ e que, pelo Teorema Limite Central, temos que $\bar{X} \sim N(p, p(1-p)/n)$.

   a) Queremos obter o tamanho da amostra para que o erro da estimativa seja no máximo 0.06 com probabilidade 0.94. Isto é,

   $$\mathbb{P}(\mid \widehat{p} - p \mid \leq 0.06) = 0.94.$$

   Se n for grande o suficiente podemos aproximar a distribuição de $\widehat{p}$ utilizando o Teorema Limite Central, de modo que

   $$\widehat{p} \sim N\left(\frac{p(1-p)}{n}\right).$$

   Então,

   $$\begin{aligned} \mathbb{P}(\mid \widehat{p} - p \mid \leq 0.06) &= 0.94 \\ \Rightarrow \mathbb{P}\left(\frac{-0.06}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{0.06}{\sqrt{\frac{p(1-p)}{n}}}\right) &= 0.94 \\ \Rightarrow \mathbb{P}\left(\frac{-0.06}{\sqrt{\frac{p(1-p)}{n}}} \leq Z \leq \frac{0.06}{\sqrt{\frac{p(1-p)}{n}}}\right) &= 0.94. \end{aligned}$$

   Logo,

   $$\frac{0.06}{\sqrt{\frac{p(1-p)}{n}}} = 1.88,$$

e

$$n = \frac{p(1-p)1.88^2}{0.06^2} = p(1-p)\cdot 981.78.$$

Como p é desconhecido, substituimos $p(1-p)$ por seu valor máximo, 0.25. Note que o gráfico de $g(p) = p(1-p)$ é uma parábola com concavidade para baixo e o máximo ocorre em $p = 0.5$. Assim, tomamos

$$n = \frac{0.25\cdot 1.88^2}{0.06^2} = 245.44.$$

Assim, o tamanho amostral para que o erro da estimativa seja no máximo 0.06 com probabilidade 0.94 é 246.

b) Sabendo que $p \leq 0.35$, temos que o valor máximo de $p(1-p)$ é $0.35\cdot 0.65 = 0.23$. Lembre que $g(p)$ é crescente para $p \leq 0.5$. Assim, tomamos

$$n = \frac{0.23\cdot 1.88^2}{0.06^2} = 225.81.$$

Logo, é possível considerar em a) uma amostra de tamanho menor, basta tomar $n = 226$.

c) Uma estimativa pontual para p é $\widehat{p} = 36/180 = 0.2$. Um intervalo de confiança para p com 95% de confiança é dado por

$$\left[0.2 - 1.96\sqrt{\frac{0.2\cdot 0.8}{180}}, 0.2 + 1.96\sqrt{\frac{0.2\cdot 0.8}{180}}\right] = [0.14, 0.26].$$

4. No problema considerado na questão anterior, suponha que se sabe que 20% das árvores estão contaminadas pela praga. Com o intuito de diminuir a proporção dessa contaminação, o biólogo da Secretaria do Meio Ambiente implantou algumas medidas de tratamento e proteção da área verde da cidade. Após seis meses, para verificar se as medidas causaram o efeito desejado, foram selecionadas ao acaso 100 árvores da cidade e foi feita a avaliação da saúde dessas árvores.

   a) (1 ponto) Formule o problema como um teste de hipóteses, enunciando a hipótese nula e a hipótese alternativa. Muita atenção aqui, porque se você errar esse item errará o seguinte.

   b) (1,5 ponto) Se na amostra foram observadas 13 árvores contaminadas, qual é a conclusão do biólogo? Dê sua resposta com base na construção de **região crítica** de um teste de hipóteses. Use um nível de significância $\alpha = 10\%$.

Solução:

Considere a variável aleatória X que assume o valor 1 se uma árvore selecionada ao acaso é contaminada e 0 caso contrário. Note que $X \sim Bernoulli(p)$. Consideremos a construção do teste seguindo os 5 passos.

a) *Passo 1*: Note que o biólogo quer verificar se a proporção de árvores contaminadas diminuiu, então, consideremos as seguintes hipóteses:

$$\mathcal{H}_0 : p = 0.2 \quad vs \quad \mathcal{H}_1 : p < 0.2.$$

b) *Passo 2*: Vamos usar a estatística $\widehat{p}$, que representa a proporção de 100 árvores da área florestal que estão contaminadas. Ainda, sabemos que, pelo Teorema Limite Central, $\widehat{p} \sim N(p, p(1-p)/100)$, assumindo que 100 é um tamanho amostral suficientemente grande.

Passo 3: Fixe $\alpha = 10\%$ e, sob a hipótese nula, $\widehat{p} \sim N(0.2, 0.2 \cdot 0.8/100)$. Queremos encontrar uma região crítica da forma $RC = \{\widehat{p} \leq c\}$, em que c é tal que $\mathbb{P}(erroI) = 0.1$. Dessa forma, temos que

$$\mathbb{P}(\widehat{p} \leq c \mid p = 0.2) = 0.1 \Rightarrow \mathbb{P}(Z \leq \frac{c - 0.2}{\sqrt{0.16/100}}) = 0.1 \Rightarrow \frac{c - 0.2}{\sqrt{0.16/100}} = -1.28$$

$$\Rightarrow c - 0.2 = -0.05 \Rightarrow c = 0.15.$$

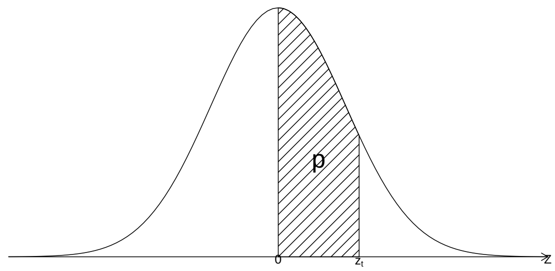
Assim, a região crítica é

$$RC = \{\widehat{p} \leq 0.15\}.$$

Passo 4: A proporção amostral é $\widehat{p} = 13/100 = 0.13$.

Passo 5: Como o valor observado de $\widehat{p}$ pertence à região crítica, temos evidências para rejeitar a hipótese nula, considerando $\alpha = 10\%$. Isto é, temos evidências de que as medidas implantadas fizeram com que a proporção de árvores contaminadas diminuísse.

Distribuição Normal



	0	1	2	3	4	5	6	7	8	9
0,0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0,1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0,2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0,3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0,4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0,5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0,6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0,7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0,8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0,9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1,0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1,1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1,2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1,3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41309	0.41466	0.41621	0.41774
1,4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1,5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1,6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1,7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1,8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1,9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2,0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2,1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2,2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2,3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2,4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2,5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2,6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2,7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2,8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2,9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3,0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900
3,1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3,2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950
3,3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965
3,4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976
3,5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983
3,6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989
3,7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992
3,8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995
3,9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997

Tabela 1: Probabilidades $p = P[0 \leq Z \leq Z_t]$ da Distribuição Normal padrão com valores de Z_t dados nas margens da tabela