

Curso : Probabilidad y Estadística
Sigla : EYP1113
Profesores : Ricardo Aravena, Ricardo Olea y Claudia Wehrhahn
Ayudantes : Daniela Castro, Erwin Agüero y Carlos Cayuman.

Examen

Problema 1 [30 %]

Sean X_1, X_2, \dots, X_n una muestra aleatoria proveniente de una población Poisson de parámetro ν .

- (a) [15 %] Obtenga el estimador máximo verosímil de ν y proponga una distribución aproximada.
(b) [15 %] En base al resultado obtenido en (a) y a las realizaciones logradas por una investigación:

$$n = 112 \quad \sum_{i=1}^n x_i = 1124 \quad \sum_{i=1}^n x_i^2 = 12632 \quad \sum_{i=1}^n \ln(x_i) = 251.6107$$

¿cuál debería ser la confianza para que el error de estimación fuese igual ± 0.3 ?

Solución

- (a) La función de verosimilitud es

$$\begin{aligned} L(X_1, \dots, X_n, \nu) &= \prod_{i=1}^n \frac{e^{-\nu} \nu^{X_i}}{X_i!} \\ &= e^{-n\nu} \nu^{\sum_{i=1}^n X_i} \prod_{i=1}^n \frac{1}{X_i!} \quad [1\%] \end{aligned}$$

Luego,

$$\ln(L(X_1, \dots, X_n, \nu)) = -n\nu + \ln(\nu) \sum_{i=1}^n X_i - \sum_{i=1}^n \ln(X_i!).$$

Para encontrar el máximo,

$$\frac{\partial \ln L}{\partial \nu} = -n + \frac{1}{\nu} \sum_{i=1}^n X_i$$

de donde se obtiene que

$$\hat{\nu} = \bar{X}_n.$$

Para verificar que es máximo

$$\begin{aligned} \frac{\partial^2 \ln(L)}{\partial \nu^2} \Big|_{\nu=\hat{\nu}} &= -\frac{1}{\nu^2} \sum_{i=1}^n X_i \Big|_{\nu=\hat{\nu}} \\ &= -\frac{n^2}{\sum_{i=1}^n X_i} < 0 \quad [1\%] \end{aligned}$$

Por lo tanto el estimador de máxima verosimilitud es

$$\hat{\nu} = \bar{X}_n. \quad [4\%]$$

Por propiedades del EMV, tenemos que

$$\hat{\nu} \sim \text{Normal}(\nu, \sqrt{1/I_n(\nu)}), \quad [2\%]$$

donde

$$\begin{aligned} I_n(\nu) &= -E\left(\frac{\partial^2 \ln(L)}{\partial \nu^2}\right) \quad [1\%] \\ &= -E\left(-\frac{1}{\nu^2} \sum_{i=1}^n X_i\right) \\ &= \frac{n}{\nu} \quad [2\%] \end{aligned}$$

Por lo tanto

$$\hat{\nu} = \bar{X}_n \sim \text{Normal}(\nu, \sqrt{\nu/n}) \quad [4\%]$$

(b) Tenemos que Un estimador de $\sqrt{\nu/n}$ es $\sqrt{\bar{X}_n/n}$. Así, se tiene que

$$\frac{\bar{X}_n - \lambda}{\sqrt{\bar{X}_n/n}} \sim \text{Normal}(0, 1) \quad [4\%]$$

Luego, el error de estimación está dado por

$$w = k_{1-\alpha/2} \sqrt{\frac{\bar{X}_n}{n}}, \quad [2\%]$$

donde $k_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, con $\Phi()$ la función de distribución acumulada de una distribución $N(0, 1)$.

A partir de la información entregada, se tiene que

$$\begin{aligned} 0.3 &= k_{1-\alpha/2} \sqrt{\frac{\bar{x}_n}{n}} \quad [3\%] \\ \Rightarrow k_{1-\alpha/2} &= 0.3 \sqrt{\frac{n}{\bar{x}_n}} \\ \Rightarrow k_{1-\alpha/2} &= 0.3 \sqrt{\frac{112}{1124/112}} \\ \Rightarrow 1 - \alpha/2 &= \Phi(1.0022) \quad [3\%] \\ \Rightarrow \alpha &= 2(1 - 0.8418) \\ \Rightarrow \alpha &= 0.3164 \\ \Rightarrow 1 - \alpha &= 0.6836 \quad [3\%] \end{aligned}$$

Luego, el intervalo debe ser de 68.38 % de confianza para tener un error de estimación de ± 0.3

Problema 2 [40 %]

El consumo de azúcar, puede ser nocivo para la salud, por ejemplo puede llevar a una persona a padecer diabetes. Una encuesta de salud realizada por la universidad informo los siguientes niveles de azúcar en la sangre (mg/dL) en los adultos chilenos mayores de 15 años:

$$n = 4607 \quad \frac{1}{n} \sum_{i=1}^n x_i = 90.97352 \quad \frac{1}{n} \sum_{i=1}^n x_i^2 = 8428.722 \quad \frac{1}{n} \sum_{i=1}^n \ln(x_i) = 4.502108 \quad \frac{1}{n} \sum_{i=1}^n [\ln(x_i)]^2 = 20.28529$$

Mientras que una clasificación por intervalo fue la siguiente:

Intervalo	Observado
50 - 70	52
70 - 90	2616
90 - 110	1611
110 - 130	258
130 - 150	70

Entre una distribución Normal y Log-Normal, ¿cual prefiere? Justifique.

Solución

Los estimadores máximo verosímiles están dados por:

- Normal(μ , σ):

$$[2\%] \quad \hat{\mu} = 90.97352; \quad \hat{\sigma} = \sqrt{8428.722 - 90.97352^2} = 12.35074 \quad [4\%]$$

- Log-Normal(λ , ζ):

$$[2\%] \quad \hat{\lambda} = 4.502108; \quad \hat{\zeta} = \sqrt{20.28529 - 4.502108^2} = 0.1277245 \quad [4\%]$$

Luego,

Intervalo	O	p (Normal)	p (Log-Normal)	E_{Normal}	$E_{\text{Log-Normal}}$	$(O - E_{\text{Normal}})^2 / E_{\text{Normal}}$	$(O - E_{\text{Log-Normal}})^2 / E_{\text{Log-Normal}}$
50 - 70	52	0.0447	0,0235	205,9329	108,2645	115,0634	29.2404
70 - 90	2616	0.4238	0,4693	1952,4466	2162,0651	225,5135	95.3056
90 - 110	1611	0.4697	0,4470	2163,9079	2059,3290	141,2755	97.6041
110 - 130	258	0.0609	0,0581	280,5663	267,6667	1,8150	0.3491
130 - 150	70	0.0009	0,0021	4,1463	9,6747	1045,9228	376.1504
Σ	4607	1.0000	1.0000	4607.0000	4607.0000	1529.5903	598.6495
		[4 %]	[4 %]	[4 %]	[4 %]	[4 %]	[4 %]

Como ambos estadísticos de pruebas distribuyen $\chi^2(2)$, entonces es claro que un mejor ajuste se logra con la distribución Log-Normal. [4 %]

Problema 3 [30 %]

Sean X e Y variables aleatorias continuas con densidad conjunta dada por

$$f_{X,Y}(x,y) = \frac{\nu^k}{\Gamma(k)} y^{k-3} e^{-\nu y}, \quad 0 \leq x \leq y^2, \quad y \geq 0$$

con $k > 0$ y $\nu > 0$. Obtenga $\mathbf{E}(X)$ y $\mathbf{Var}(X)$.

Solución

Se pide

$$\mathbf{E}(X) = \mathbf{E}[\mathbf{E}(X | Y)] \quad \text{y} \quad \mathbf{Var}(X) = \mathbf{Var}[\mathbf{E}(X | Y)] + \mathbf{E}[\mathbf{Var}(X | Y)] \quad [2 \%]$$

Notemos que

$$f_Y(y) = \int_0^{y^2} \frac{\nu^k}{\Gamma(k)} y^{k-3} e^{-\nu y} dx = \frac{\nu^k}{\Gamma(k)} y^{k-1} e^{-\nu y}, \quad y \geq 0 \quad [2 \%]$$

es decir, $Y \sim \text{Gamma}(k, \nu)$. [2 %]

Por otra parte

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{y^2}, \quad 0 \leq x \leq y^2 \quad [2 \%]$$

es decir, $X | Y = y \sim \text{Uniforme}(0, y^2)$. [2 %]

Reemplazando

$$\mathbf{E}(X) = \frac{1}{2} \cdot \mathbf{E}(Y^2) \quad [2 \%] \tag{1}$$

y

$$\mathbf{Var}(X) = \frac{1}{4} \cdot \mathbf{Var}(Y^2) + \frac{1}{12} \cdot \mathbf{E}(Y^4) \quad [2 \%] \tag{2}$$

Del formulario, tenemos que la función generadora de momentos de una variable aleatoria $\text{Gamma}(k, \nu)$ es:

$$M_Y(t) = \left(\frac{\nu}{\nu - t} \right)^k \quad [2 \%]$$

A partir de esta se tiene que

$$\mathbf{E}(Y) = \frac{k}{\nu}, \quad \mathbf{E}(Y^2) = \frac{k(k+1)}{\nu^2}, \quad \mathbf{E}(Y^3) = \frac{k(k+1)(k+2)}{\nu^3}, \quad \mathbf{E}(Y^4) = \frac{k(k+1)(k+2)(k+3)}{\nu^4}$$

[6 %]

Por lo tanto

$$\mathbf{E}(X) = \frac{k(k+1)}{2\nu} \quad [2 \%]$$

y

$$\mathbf{Var}(X) = \frac{k(k+1)(5k^2 + 37k + 48)}{12\nu^4} \quad [4 \%]$$

Formulario

- Sea X_1, \dots, X_n una muestra aleatoria independiente e idénticamente distribuida con función de probabilidad p_X o de densidad f_X , determinada por un parámetro θ . Si $\hat{\theta}$ es el estimador máximo verosímil del parámetro θ , entonces:

- $\mathbf{E}(\hat{\theta}) \rightarrow \theta$, cuando $n \rightarrow \infty$.

- $\mathbf{Var}(\hat{\theta}) = \frac{1}{I_n(\theta)}$, con $I_n(\theta) = -\mathbf{E} \left[\frac{\partial^2}{\partial \theta^2} \ln L(\theta) \right]$.

- $\hat{\theta} \sim \text{Normal} \left(\theta, \sqrt{\frac{1}{I_n(\theta)}} \right)$, cuando $n \rightarrow \infty$.

- El estimador máximo verosímil de $g(\theta)$ es $g(\hat{\theta})$, cuya varianza está dada por: $\mathbf{Var}[g(\hat{\theta})] = \frac{[g'(\theta)]^2}{I_n(\theta)}$.

- Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas $\text{Normal}(\mu, \sigma)$, entonces

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0, 1), \quad \frac{\bar{X}_n - \mu}{s/\sqrt{n}} \sim \text{t-student}(n-1), \quad \frac{s^2(n-1)}{\sigma^2} \sim \chi_{n-1}^2$$

$$\text{con } s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

- **Propiedades función $\Gamma(\cdot)$**

$$(1) \quad \Gamma(k) = \int_0^\infty u^{k-1} e^{-u} du; \quad (2) \quad \Gamma(a+1) = a \Gamma(a);$$

$$(3) \quad \Gamma(n+1) = n!, \quad \text{si } n \in \mathbb{N}_0; \quad (4) \quad \Gamma(1/2) = \sqrt{\pi}$$

- **Propiedades función $B(\cdot, \cdot)$**

$$(1) \quad B(q, r) = \int_0^1 x^{q-1} (1-x)^{r-1} dx; \quad (2) \quad B(q, r) = \frac{\Gamma(q) \Gamma(r)}{\Gamma(q+r)}$$

- **Igualdades**

$$\sum_{k=0}^n \binom{n}{k} a^k b^{n-k} = (a+b)^n, \quad \sum_{k=x}^{\infty} \phi^k = \frac{\phi^x}{1-\phi} \quad \text{si } |\phi| < 1, \quad \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \exp(\lambda)$$

Distribuciones

Distribución	Densidad de Probabilidad	Θ_X	Parámetros	Esperanza y Varianza
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$	$x = 0, \dots, n$	n, p	$\mu_X = n p$ $\sigma_X^2 = n p (1-p)$ $M(t) = [p e^t + (1-p)]^n, \quad t \in \mathbb{R}$
Geométrica	$p (1-p)^{x-1}$	$x = 1, 2, \dots$	p	$\mu_X = 1/p$ $\sigma_X^2 = (1-p)/p^2$ $M(t) = p e^t / [1 - (1-p) e^t], \quad t < -\ln(1-p)$
Binomial-Negativa	$\binom{x-1}{r-1} p^r (1-p)^{x-r}$	$x = r, r+1, \dots$	r, p	$\mu_X = r/p$ $\sigma_X^2 = r(1-p)/p^2$ $M(t) = \{p e^t / [1 - (1-p) e^t]\}^r, \quad t < -\ln(1-p)$
Poisson	$\frac{(\nu t)^x e^{-\nu t}}{x!}$	$x = 0, 1, \dots$	ν	$\mu_X = \nu t$ $\sigma_X^2 = \nu t$ $M(t) = \exp\left[\lambda (e^t - 1)\right], \quad t \in \mathbb{R}$
Exponencial	$\nu e^{-\nu x}$	$x \geq 0$	ν	$\mu_X = 1/\nu$ $\sigma_X^2 = 1/\nu^2$ $M(t) = \nu/(\nu - t), \quad t < \nu$
Gamma	$\frac{\nu^k}{\Gamma(k)} x^{k-1} e^{-\nu x}$	$x \geq 0$	k, ν	$\mu_X = k/\nu$ $\sigma_X^2 = k/\nu^2$ $M(t) = [\nu/(\nu - t)]^k, \quad t < \nu$
Normal	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$	$-\infty < x < \infty$	μ, σ	$\mu_X = \mu$ $\sigma_X^2 = \sigma^2$ $M(t) = \exp(\mu t + \sigma^2 t^2/2), \quad t \in \mathbb{R}$
Log-Normal	$\frac{1}{\sqrt{2\pi}(\zeta x)} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\zeta}\right)^2\right]$	$x \geq 0$	λ, ζ	$\mu_X = \exp\left(\lambda + \frac{1}{2}\zeta^2\right)$ $\sigma_X^2 = \mu_X^2 (e^{\zeta^2} - 1)$ $E(X^r) = e^{r\lambda} M_Z(r\zeta), \text{ con } Z \sim \text{Normal}(0,1)$
Uniforme	$\frac{1}{(b-a)}$	$a \leq x \leq b$	a, b	$\mu_X = (a+b)/2$ $\sigma_X^2 = (b-a)^2/12$ $M(t) = [e^t b - e^t a]/[t(b-a)], \quad t \in \mathbb{R}$
Beta	$\frac{1}{B(q, r)} \frac{(x-a)^{q-1} (b-x)^{r-1}}{(b-a)^{q+r-1}}$	$a \leq x \leq b$	q, r	$\mu_X = a + \frac{q}{q+r} (b-a)$ $\sigma_X^2 = \frac{q r (b-a)^2}{(q+r)^2 (q+r+1)}$
Hipergeométrica	$\frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}$	$\max\{0, n+m-N\} \leq x \leq \min\{n, m\}$	N, m, n	$\mu_X = n \frac{m}{N}$ $\sigma_X^2 = \left(\frac{N-n}{N-1}\right) n \frac{m}{N} \left(1 - \frac{m}{N}\right)$

Tablas de Percentiles p

Distribución Normal Estándar k_p											Distribución t-student $t_p(\nu)$				
k_p	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	ν	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	1	3.078	6.314	12.706	31.821
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	2	1.886	2.920	4.303	6.965
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	3	1.638	2.353	3.182	4.541
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	1.533	2.132	2.776	3.747
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	5	1.476	2.015	2.571	3.365
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	6	1.440	1.943	2.447	3.143
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	7	1.415	1.895	2.365	2.998
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	8	1.397	1.860	2.306	2.896
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	9	1.383	1.833	2.262	2.821
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	10	1.372	1.812	2.228	2.764
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	11	1.363	1.796	2.201	2.718
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	12	1.356	1.782	2.179	2.681
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	13	1.350	1.771	2.160	2.650
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	14	1.345	1.761	2.145	2.624
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	15	1.341	1.753	2.131	2.602
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	16	1.337	1.746	2.120	2.583
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	17	1.333	1.740	2.110	2.567
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	18	1.330	1.734	2.101	2.552
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	19	1.328	1.729	2.093	2.539
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	20	1.325	1.725	2.086	2.528
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	21	1.323	1.721	2.080	2.518
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	22	1.321	1.717	2.074	2.508
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	23	1.319	1.714	2.069	2.500
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	24	1.318	1.711	2.064	2.492
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	25	1.316	1.708	2.060	2.485
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	26	1.315	1.706	2.056	2.479
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	27	1.314	1.703	2.052	2.473
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	28	1.313	1.701	2.048	2.467
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	29	1.311	1.699	2.045	2.462
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	30	1.310	1.697	2.042	2.457
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990	∞	1.282	1.645	1.960	2.326

Distribución Chi-Cuadrado $c_p(\nu)$								
ν	$c_{0.025}$	$c_{0.05}$	$c_{0.10}$	$c_{0.90}$	$c_{0.95}$	$c_{0.975}$	$c_{0.99}$	$c_{0.995}$
1	0.00	0.00	0.02	2.71	3.84	5.02	6.63	7.88
2	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93