Linköpings Universitet
IDA/Statistik
LH

Exam in Statistical Methods, 2015-05-29

Time allowed: kl: 8-12

Allowed aids: Calculator. One handwritten A4 paper (both sides) with the students own notes.

Assisting teacher: Lotta Hallberg

Grades: A=19-20 points, B=17-18p, C=14-16p, D=12-13p, E=10-11p

Provide a detailed report that shows motivation of the results.

1

Let the random variable (Y_1, Y_2) have density function $f(y_1, y_2) = e^{-y_1}$ if $0 < y_2 \le y_1 < \infty$ and 0 elsewhere.

a) Determine if Y_1 and Y_2 are independent using the definition independent random variables. 1p

b) Calculate $P(Y_2 < 1)$ 2p

c) Determine by either calculating or by recognition of the distribution, $E[Y_1]$ 3p

2

Suppose that the posterior distribution of the mean μ (in a normal distribution) is normally distributed with mean=10 and standard deviation=2.

Calculate a 90% credibility interval for μ. 3p

3

Let the random variable Y be exponentially distributed with density function $f(y) = \lambda e^{-\lambda y}$ if $y > 0$ and 0 elsewhere.

You have 20 observations of Y and the sum of them is 3,528.

a) Estimate λ with the method of moments. 1p

b) Estimate λ with the maximum likelihood method. 2p

c) Let $E[Y] = \mu$. Show that $\bar{Y}$ is an unbiased and consistent estimate of μ. 2p

d) Estimate λ with the posterior Bayes estimate when the prior density of λ is $g(\lambda) = 5e^{-5\lambda}$ if $\lambda > 0$ and 0 elsewhere. 3p

4

As part of a test of solar energy, you measure the total heat flux from 29 homes. You wish to examine whether total heat flux can be predicted by the position of the focal points in the east, south, and north directions.

A multiple regression model was fit to data. The result is:

$$\hat{\beta} = \begin{pmatrix} 389,2 \\ 2,12 \\ 5,318 \\ -24,13 \end{pmatrix} \begin{array}{l} \text{int} \\ \text{east} \\ \text{south} \\ \text{nort} \end{array} \quad (\mathbf{X'X})^{-1} = \begin{array}{c} \begin{array}{cccc} \text{int} & \text{east} & \text{south} & \text{north} \end{array} \\ \begin{pmatrix} 59,0941 & -0,887432 & -0,509889 & -0,586623 \\ -0,8874 & 0,019952 & 0,004903 & 0,000776 \\ -0,5099 & 0,004903 & 0,012543 & -0,006458 \\ -0,5866 & 0,000776 & -0,006458 & 0,047230 \end{pmatrix} \end{array}$$

$S = 8,59782$

Calculate a 95% confidence interval for β_1 using the formula $\boldsymbol{a}'\hat{\beta} \pm t_{\alpha/2} S\sqrt{\boldsymbol{a}'(\mathbf{X'X})^{-1}\boldsymbol{a}}$

Test, using the interval, the hypothesis: $\beta_1 = 0$. 3p 2

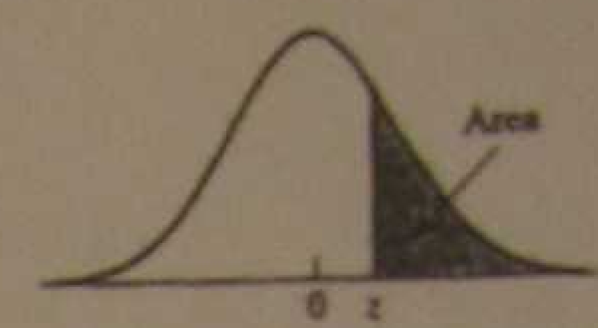
Continuous Distributions

Distribution	Probability Function	Mean	Variance	Moment-Generating Function
Uniform	$f(y) = \frac{1}{\theta_2 - \theta_1};\ \theta_1 \le y \le \theta_2$	$\frac{\theta_1 + \theta_2}{2}$	$\frac{(\theta_2 - \theta_1)^2}{12}$	$\frac{e^{t\theta_2} - e^{t\theta_1}}{t(\theta_2 - \theta_1)}$
Normal	$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\left(\frac{1}{2\sigma^2}\right)(y-\mu)^2\right]$ $-\infty < y < +\infty$	μ	σ^2	$\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$
Exponential	$f(y) = \frac{1}{\beta} e^{-y/\beta};\quad \beta > 0$ $0 < y < \infty$	β	β^2	$(1 - \beta t)^{-1}$
Gamma	$f(y) = \left[\frac{1}{\Gamma(\alpha)\beta^\alpha}\right] y^{\alpha-1} e^{-y/\beta};$ $0 < y < \infty$	$\alpha\beta$	$\alpha\beta^2$	$(1 - \beta t)^{-\alpha}$
Chi-square	$f(y) = \frac{(y)^{(v/2)-1} e^{-y/2}}{2^{v/2}\Gamma(v/2)};$ $y^2 > 0$	v	$2v$	$(1-2t)^{-v/2}$
Beta	$f(y) = \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right] y^{\alpha-1}(1-y)^{\beta-1};$ $0 < y < 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	does not exist in closed form

Discrete Distributions

Distribution	Probability Function	Mean	Variance	Moment-Generating Function
Binomial	$p(y) = \binom{n}{y} p^y (1-p)^{n-y};$ $y = 0, 1, \ldots, n$	np	$np(1-p)$	$[pe^t + (1-p)]^n$
Geometric	$p(y) = p(1-p)^{y-1};$ $y = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Hypergeometric	$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}};$ $y = 0, 1, \ldots, n$ if $n \le r$, $y = 0, 1, \ldots, r$ if $n > r$	$\frac{nr}{N}$	$n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$	
Poisson	$p(y) = \frac{\lambda^y e^{-\lambda}}{y!};$ $y = 0, 1, 2, \ldots$	λ	λ	$\exp[\lambda(e^t - 1)]$
Negative binomial	$p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r};$ $y = r, r+1, \ldots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$

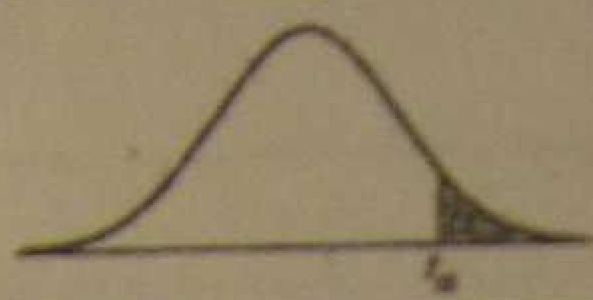
Table 4 Normal Curve Areas
Standard normal probability in right-hand tail
(for negative values of z, areas are found by symmetry)



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
	Second decimal place of z									
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000 233									
4.0	.000 031 7									
4.5	.000 003 40									
5.0	.000 000 287									

From R. E. Walpole, *Introduction to Statistics* (New York: Macmillan, 1968).

Table 5 Percentage Points of the *t* Distributions



$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	df
3.078	6.314	12.706	31.821	63.657	1
1.886	2.920	4.303	6.965	9.925	2
1.638	2.353	3.182	4.541	5.841	3
1.533	2.132	2.776	3.747	4.604	4
1.476	2.015	2.571	3.365	4.032	5
1.440	1.943	2.447	3.143	3.707	6
1.415	1.895	2.365	2.998	3.499	7
1.397	1.860	2.306	2.896	3.355	8
1.383	1.833	2.262	2.821	3.250	9
1.372	1.812	2.228	2.764	3.169	10
1.363	1.796	2.201	2.718	3.106	11
1.356	1.782	2.179	2.681	3.055	12
1.350	1.771	2.160	2.650	3.012	13
1.345	1.761	2.145	2.624	2.977	14
1.341	1.753	2.131	2.602	2.947	15
1.337	1.746	2.120	2.583	2.921	16
1.333	1.740	2.110	2.567	2.898	17
1.330	1.734	2.101	2.552	2.878	18
1.328	1.729	2.093	2.539	2.861	19
1.325	1.725	2.086	2.528	2.845	20
1.323	1.721	2.080	2.518	2.831	21
1.321	1.717	2.074	2.508	2.819	22
1.319	1.714	2.069	2.500	2.807	23
1.318	1.711	2.064	2.492	2.797	24
1.316	1.708	2.060	2.485	2.787	25
1.315	1.706	2.056	2.479	2.779	26
1.314	1.703	2.052	2.473	2.771	27
1.313	1.701	2.048	2.467	2.763	28
1.311	1.699	2.045	2.462	2.756	29
1.282	1.645	1.960	2.326	2.576	inf.

From "Table of Percentage Points of the *t*-Distribution." Computed by Maxine Merrington, *Biometrika*, Vol. 32 (1941), p. 300.

STATISTICAL METHODS EXAM 29-5-2015

① $f(y_1, y_2) = e^{-y_1}$ if $0 < y_2 \leq y_1 < \infty$ and 0 elsewhere

ⓐ if $f(y_1, y_2) = f(y_1) \cdot f(y_2)$ they are independent

$$f(y_1) = \int_0^{y_1} e^{-y_1} dy_2 = e^{-y_1} [y_2]_0^{y_1} = \underline{y_1 e^{-y_1} = f(y_1)}$$

$$f(y_2) = \int_{y_2}^{\infty} e^{-y_1} dy_1 = \left[-e^{-y_1}\right]_{y_2}^{\infty} = 0 + e^{-y_2} = \underline{f(y_2) = e^{-y_2}}$$

$f(y_1, y_2) \neq f(y_1) f(y_2)$ so they are not independent

ⓑ $$P(Y_2 < 1) = F_{Y_2}(1) = \int_0^1 e^{-y_2} dy_2 = \left[-e^{-y_2}\right]_0^1 = -\frac{1}{e} + 1 = \underline{1 - \frac{1}{e}}$$

ⓒ $E[Y_1]$?

$f(y_1) = y_1 e^{-y_1}$ which is also Gamma ($\alpha = 2$, $\beta = 1$)

$$= f(y_1) = \left[\frac{1}{\underbrace{\Gamma(2)}_{1} \underbrace{1^2}_{1}}\right] y_1^{2-1} e^{-y_1/1}$$

And since the expectation of Gamma is $\alpha\beta$ then $E[Y_1] = 2 \cdot 1 = \underline{2}$

② Posterior of μ is $N(10,4)$.

Since $Z = \frac{\bar{x}-\mu}{\sigma} \sim N(0,1)$ we can create this interval:

$P\left(Z_{-\alpha/2} \leq \frac{\mu - 10}{2} \leq Z_{\alpha/2}\right) = 1-\alpha$, with $\alpha = 0.1$ and $Z_{-\alpha/2} = -1.64$, $Z_{\alpha/2} = 1.64$

Replacing and operating we have:

$P(-1{,}64 \cdot 2 + 10 \leq \mu \leq 1.64 \cdot 2 + 10) = 0.9 \Rightarrow \underline{P(6{,}72 \leq \mu \leq 13{,}28) = 0.9}$

③ $f(y) = \lambda e^{-\lambda y}$, $y > 0$, 0 elsewhere, and $n = 20$, $\Sigma y = 3{,}528$

ⓐ Estimate λ by method of moments:

$\mu_1' = E[Y] = \frac{1}{\lambda}$

$m_1' = \frac{\Sigma y}{n}$

$\Rightarrow \mu_1' = m_1' \Rightarrow \frac{1}{\lambda} = \frac{\Sigma y}{n} \Rightarrow \hat{\lambda} = \frac{n}{\Sigma y} = \frac{20}{3{,}528} \Rightarrow \underline{\hat{\lambda} = 5{,}67}$

ⓑ Estimate λ by maximum likelihood

$\mathcal{L}(y|\lambda) = \prod_{i=1}^{n} f(y) = \lambda^n e^{-\lambda \Sigma y} \Rightarrow$ taking logarithms we have:

$\Rightarrow \log \mathcal{L}(y|\lambda) = n \log \lambda + \lambda \Sigma y$; now taking derivative with respect to λ

and equaling to 0 we have $\Rightarrow \frac{n}{\lambda} - \Sigma y = 0 \Rightarrow \hat{\lambda} = \frac{n}{\Sigma y} \Rightarrow \underline{\hat{\lambda} = 5{,}67}$

ⓒ Let $E[y] = \mu$. Show that $\bar{y}$ is unbiased and consistent estimate of μ

* Unbiased: The mean of the estimate is equal to the parameter

$$E[\bar{y}] = E\left[\frac{1}{n}\sum y\right] = \frac{1}{n}\sum E[y] = \frac{1}{\cancel{n}}\cancel{n}\mu = \underline{\mu} \rightarrow \text{unbiased}$$

* Consistent: The variance of the estimator goes to 0 as n goes to ∞

$$V[\bar{y}] = V\left[\frac{1}{n}\sum y\right] = \frac{1}{n^2}\sum V[y] = \frac{1}{n^2}\cancel{n}\sigma^2 = \underline{\frac{\sigma^2}{n}} \rightarrow 0 \text{ as } n \rightarrow \infty$$

↳ consistent!

ⓓ Estimate λ with posterior Bayes estimate when the prior of λ is $g(\lambda) = 5e^{-5\lambda}$

① We have to get the posterior of λ; Posterior = prior x likelihood

$$\mathcal{L}(y|\lambda) = \lambda^n e^{-\lambda\sum y} \text{ and } g(\lambda) = 5e^{-5\lambda} \text{ so}$$

$$g^*(\lambda|y) \propto \mathcal{L}(y|\lambda)\,p(\lambda) = 5\lambda^n e^{-\lambda\sum y} e^{-5\lambda} = \underline{5\lambda^n e^{-\lambda(\sum y+5)}}$$

Posterior density

② The Bayes estimate $\hat{\lambda}$ is $E\left(g^*(\lambda|y)\right)$

If we replace n and $\sum y$ in the posterior we get

$$g^*(\lambda|y) = 5\lambda^{20} e^{-8{,}53\lambda}, \text{ which is a Gamma with } \begin{cases}\alpha = 21 \\ \beta = \frac{1}{8{,}53}\end{cases}$$

And since $E[\Gamma(\alpha,\beta)] = \alpha\beta$ then $\hat{\lambda} = \frac{21}{8{,}53} = 18{,}3$

(4) Calculate 95% c.i. for β_1 using $a'\hat{\beta} \pm t_{\alpha/2} S \sqrt{a'(x'x)^{-1}a}$

Test $\beta_1 = 0$

(a) We have that $n = 29$, $S = 8{,}59782$

The regression formula is $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$, i.e.

Total Heat Flux = 389,2 + $\underbrace{2{,}12}_{\hat{\beta}_1}$ EAST + 5,38 SOUTH − 24.13 NORTH

Confidence interval for $\hat{\beta}_1$ is $a'\hat{\beta} \pm t_{\alpha/2} S \sqrt{a'(x'x)^{-1}a}$

Here from the table we know $t_{.95}(24) = 2.064$; and $a' = (0\ 1\ 0\ 0)$

$a'(x'x)^{-1}a = \quad = 0.019952$ (element 2,1 of $(XX)^{-1}$)

So the confidence interval is $2{,}12 \pm 2064 \cdot 8{,}59 \sqrt{0.019952}$

$= 2{,}12 \pm 0.05$

(b) $\beta_1 = 0$? The value 0 is outside the acceptance region given by, so we reject the hypothesis $H_0: \beta_1 = 0$