Statistics

Homework 4 (Due May 7, 2021)

Joint distributions

1. Random variables $X_1, X_2, \cdots, X_n$ have a common density $f_X(\cdot)$ with variance σ^2.

   Let $\overline{X_n} = (X_1 + X_2 + \cdots + X_n)/n$. Prove that

   $$Var\left[\overline{X_n}\right] = \frac{\sigma^2}{n} + \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov(X_i, X_j).$$

2. Two insect species (predator and prey) coexist in a study area. The numbers of insects of the two species found in the study area can be described by a bivariate normal distribution. Let X and Y represent the numbers of the predator and prey species, respectively. A long term ecological survey found that

   $P(Y > 86 \mid X = 80) = 0.04779,\ \ P(Y < 50 \mid X = 80) = 0.09121$

   $P(Y > 55 \mid X = 150) = 0.20233,\ \ \rho_{XY} = -0.6$ (correlation coefficient of *X* and *Y*).

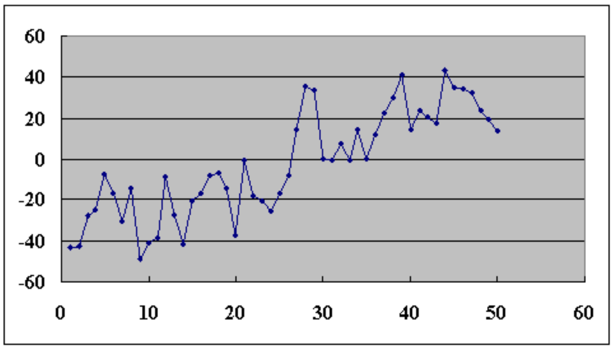
   (1) Find the standard deviations of *X* and *Y*, respectively.

   (2) What should the maximum number of the predator species be in order to achieve a higher than 75% chance for the number of prey species to be more than 70?

3. The data in the following table is a subseries of an AR(2) process –

   $$x(t) = 0.62x(t-1) + 0.25x(t-2) + \varepsilon(t) \quad \varepsilon(t) \sim iid\ N\left(0, \sigma_\varepsilon^2 = 225\right)$$

t	1	2	3	4	5	6	7	8	9	10
0	-43.5615	-42.5594	-27.5556	-24.8670	-7.6353	-16.6590	-30.3863	-13.9924	-48.8766	-40.9939
10	-38.2022	-8.6007	-26.9491	-41.7347	-20.3306	-16.6453	-8.1240	-6.7680	-14.3989	-37.0406
20	-0.8470	-18.0262	-20.5075	-25.0793	-16.8221	-7.7765	14.4545	35.3056	33.6065	-0.1143
30	-0.5319	7.5514	-0.4998	14.1964	-0.1449	11.8305	22.5281	29.9923	40.7217	14.2982
40	23.3038	20.4037	17.2357	43.1575	34.5701	33.7827	31.9995	23.4265	18.9954	13.7399

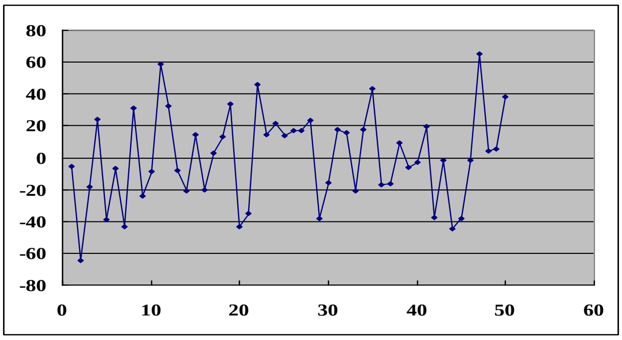


   (1) Calculate the mean and standard deviation of $x(t),\ t = 1, 2, \cdots, 50$.

   (2) The theoretical values of mean and standard deviation of X are respectively 0 and 27.5318. Compare these theoretical values against their estimates obtained in (1).

4. The following is a random sample from a normal distribution of mean 0 and standard deviation 27.5318.

5.	1	2	3	4	5	6	7	8	9	10
0	-5.407	-64.317	-18.434	24.298	-38.989	-6.800	-43.550	31.181	-23.789	-8.363
10	59.050	32.460	-8.182	-20.621	14.489	-20.165	2.946	13.432	33.898	-43.162
20	-35.052	46.106	14.658	21.340	13.584	16.901	16.977	23.161	-38.092	-16.026
30	17.982	15.979	-20.913	17.966	43.496	-17.282	-16.295	9.455	-6.092	-3.146
40	19.561	-37.615	-1.747	-44.600	-38.186	-1.586	65.115	4.176	5.509	38.330



(1) Calculate the sample mean and sample standard deviation.

(2) Are the sample estimates of mean and standard deviation in this problem closer to the theoretical values than the estimates in Problem 3?

5. The exponential density $g(x) = \lambda e^{-\lambda x}$ with $\lambda = 1$ is everywhere higher than the standard normal density $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ for $0 \le x < +\infty$.

(1) Prove and show that the rejection method with $cg(x) = \sqrt{\frac{2e}{\pi}} e^{-x}$ can be used to generate random samples of the standard normal distribution.

(2) Develop an R code to generate 10,000 random numbers from the standard normal density $(z_1,\ z_2, \cdots,\ z_n;\ n = 10{,}000)$ using the above rejection method. Calculate the mean and standard deviation of $(z_1,\ z_2, \cdots,\ z_n;\ n = 10{,}000)$. Also, show the histogram of $(z_1,\ z_2, \cdots,\ z_n;\ n = 10{,}000)$.

> Hint:
>
> (1) Generate $X \sim Exp(1)$ and $Y \sim U\left(0, \sqrt{\frac{2e}{\pi}} e^{-x}\right)$.
>
> (2) Reject *X*, if $Y > \phi_+(X)$, where $\phi_+(x) = \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}x^2}$
>
> (3) If $Y \le \phi_+(X)$, generate $S \sim U[-1, 1]$. Then, $Z = SX$.