STAT201A: Introduction to Probability at an Advanced Level (Fall 2024)
UC Berkeley

Problem Set 5

Due: 10:00pm, Wednesday, November 20, 2024 (via Gradescope)

1. **(Multivariate normal)** Suppose $Y \sim \mathcal{N}_n(\mu, \Sigma)$ in this problem.

   (a) If a is any fixed vector in $\mathbb{R}^n$, show that

   $$\frac{a^\top (Y-\mu)}{\sqrt{a^\top \Sigma a}} \sim \mathcal{N}(0,1).$$

   (b) If A is now a random vector that is independent of Y, then show again that

   $$\frac{A^\top (Y-\mu)}{\sqrt{A^\top \Sigma A}}$$

   is distributed according to $\mathcal{N}(0,1)$ and that it is independent of A.

   (c) Using the above result, show that if $Y \sim \mathcal{N}_3(0, I_3)$, then

   $$\frac{Y_1 e^{Y_3} + Y_2 \log|Y_3|}{\sqrt{e^{2Y_3} + (\log|Y_3|)^2}} \sim \mathcal{N}(0,1).$$

   (a) Suppose $Y \sim \mathrm{N_n}(\mu, \Sigma)$ and $a \in \mathbb{R}^n$. Then, since Y is Multivariate Normal,

   $$a^\top Y \sim \mathrm{N}\left(a^\top \mu, a^\top \Sigma a\right)$$

   $$\Rightarrow a^\top Y - a^\top \mu \sim \mathrm{N}\left(0, a^\top \Sigma a\right)$$

   $$\Rightarrow \frac{a^\top (Y-\mu)}{\sqrt{a^\top \Sigma a}} \sim \mathrm{N}(0,1).$$

   (b) We note that from part (a), when a is fixed, $\frac{a^\top (Y-\mu)}{\sqrt{a^\top \Sigma a}} \sim \mathrm{N}(0,1)$. Hence, its moment generation function is given by

   $$M_{\frac{a^\top (Y-\mu)}{\sqrt{a^\top \Sigma a}}}(t) = \mathbb{E}\left[e^{t\frac{a^\top (Y-\mu)}{\sqrt{a^\top \Sigma a}}}\right] = e^{\frac{1}{2}t^2}.$$

   Now, when a is a random vector independent from Y, we have

   $$\begin{aligned} M_{\frac{a^\top (Y-\mu)}{\sqrt{a^\top \Sigma a}}}(t) &= \mathbb{E}\left[e^{t\frac{a^\top (Y-\mu)}{\sqrt{a^\top \Sigma a}}}\right] \\ &= \mathbb{E}\left(\mathbb{E}\left[e^{t\frac{a^\top (Y-\mu)}{\sqrt{a^\top \Sigma a}}} \,\middle|\, a\right]\right) \\ &= \mathbb{E}\left(e^{\frac{1}{2}t^2}\right) \\ &= e^{\frac{1}{2}t^2}. \end{aligned}$$

(c) Let $Y = (Y_1, Y_2, Y_3)^T \sim \mathrm{N}_3\left((0,0,0)^T, I\right)$, then $W = (Y_1, Y_2)^T \sim \mathrm{N}_2\left((0,0)^T, I\right)$. Letting $a = \left(e^{Y_3}, \log|Y_3|\right)$ we see that aW. Then,

$$\begin{aligned} \frac{a^\top (W-\mu)}{\sqrt{a^\top \Sigma a}} &\sim \mathrm{N}(0,1) \\ \implies \frac{a^\top W}{\sqrt{a^\top a}} &\sim \mathrm{N}(0,1) \\ \implies \frac{Y_1 e^{Y_3} + Y_2 \log|Y_3|}{\sqrt{e^{2Y_3} + (\log|Y_3|)^2}} &\sim \mathrm{N}(0,1). \end{aligned}$$

2. **(Marginally normal but not bivariate normal)** Give an example of a 2×1 random vector $Y = (Y_1, Y_2)^T$ with positive definite covariance matrix such that each Y_1 and Y_2 is standard normal but Y is not bivariate normal.
This was covered in lecture 19. See the first slide. Take $Y_1 \sim_d \mathcal{N}(0,1)$ and X a random variable taking value -1 and 1 with probability $1/2$ in each case. Then $Y_= XY_1$ is also a normal. Moreover the covariance matrix of (Y_1, Y_2) is the identity matrix, so it is positive definite. However This random vector is not bivariate normal since $Y_1 + Y_2$ is not normal. For instance, $\mathbb{P}(Y_1 + Y_2 = 0) = 1/2$.

3. **(Conditional distribution)** Consider three random variables Y_1, Y_2 and Y_3 that are independent and standard normal. Let

$$\begin{aligned} X_1 &= Y_2 + Y_3, \\ X_2 &= Y_1 + Y_3, \\ X_3 &= Y_1 + Y_2. \end{aligned}$$

Find the conditional distribution of X_1 given $X_2 = X_3 = 0$.
This problem is an application of the formulas for conditional normal random variables given in lecture 19. Since $Y \sim_d \mathcal{N}\left((0,0,0)^T, I_3\right)$, taking the matrix $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, we have that $(X_1, X_2, X_3)^T = A(Y_1, Y_2, Y_3)^T$. Hence

$$(X_1, X_2, X_3) \sim_d \mathcal{N}\left((0,0,0)^T, AI_3A^T\right).$$

Here $AI_3A^T = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$. Using the formulas given in lecture we obtain for $Z = X_1|_{X_2 = X_3 = 0}$, $\mathbb{E}[Z] = 0$ and $\mathrm{Var}(Z) = 2 - (1,1)\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}^{-1}(1,1)^T = \frac{4}{3}$.

4. **(More on jointly Gaussian distributions)** Let X and Y be independent standard normal variables.

(a) For a constant k, find $\mathbb{P}(X > kY)$.
(b) If $U = \sqrt{3}X + Y$, and $V = X - \sqrt{3}Y$, find $\mathbb{P}(U > kV)$.
(c) Find $\mathbb{P}(U^2 + V^2 < 1)$.

(d) Find the conditional distribution of X given $V = v$.

(a) We can use the radial symmetry of the joint distribution of two standard independent random variables. Since the line $x = ky$ goes through 0 it divides the plane in two sections where the total density is the same. We automatically have $P(X > kY) = 1/2$. Alternatively, $X - kY \sim N(0, 1 + k^2)$ and $P(X - kY > 0) = 1/2$.

(b) Notice that $U \sim N(0,4)$ and $V \sim N(0,4)$. Furthermore, $Cov(U,V) = Cov(\sqrt{3}X + Y, X - \sqrt{3}Y)$. Using bilinearity properties of covariance this is $\sqrt{3}Var(X) - 3Cov(X,Y) + Cov(Y,X) - \sqrt{3}Var(Y) = -2Cov(X,Y) = 0$. It follows that the joint (U,V) is uncorrelated bivariate normal and that $P(U > kV) = \frac{1}{2}$ by radial symmetry of uncorrelated bivariate normal. Alternatively, you can check that $U - kV \sim N(0, (\sqrt{3} - k)^2 + (1 - k\sqrt{3})^2)$ and thus $P(U - kV > 0) = \frac{1}{2}$.

(c) $U, V \sim iid\, N(0,4)$ so $U/2, V/2 \sim iid\, N(0,1)$. It follows that

$$\left(\frac{U}{2}\right)^2 + \left(\frac{V}{2}\right)^2 \sim Exp\left(\frac{1}{2}\right).$$

Then

$$P(U^2 + V^2 < 1) = P(\left(\frac{U}{2}\right)^2 + \left(\frac{V}{2}\right)^2 < \frac{1}{4}) = 1 - e^{-\frac{1}{2}(\frac{1}{2})^2} = 1 - e^{-\frac{1}{8}}.$$

(d) $Cov(X,V) = Cov(X, X - \sqrt{3}Y) = Var(X)$. It follows that

$$Corr(X,V) = \frac{Cov(X,V)}{SD(X)SD(V)} = \frac{Var(X)}{\sqrt{Var(X)Var(V)}} = \frac{1}{\sqrt{4}} = \frac{1}{2}.$$

Hence $(X,V) \sim BVN(0,0,1,4,\rho = \frac{1}{2})$ which implies that $(X, \frac{V}{2}) \sim BVN(0,0,1,1,\rho = \frac{1}{2})$. Hence $X|(\frac{V}{2} = \frac{v}{2}) \sim N(\rho\frac{v}{2}, 1 - \rho^2) = N(\frac{1}{4}v, \frac{3}{4})$.

5. **(Wigner's surmise)** Let $X = \begin{pmatrix} X_1 & X_3 \\ X_3 & X_2 \end{pmatrix}$ with X_1 and X_2 independent $\mathcal{N}(0,1)$ and X_3 another independent $\mathcal{N}(0,1/2)$. Let λ_1 and λ_2 be two eigenvalues of X and $s = |\lambda_1 - \lambda_2|$.

(a) Prove that $s = \sqrt{(X_1 - X_2)^2 + 4X_3^2}$.

(b) Find the density of s.

(c) Plot the density function of s. What do you observe respect to the eigenvalues of the random matrix X?

Start by noticing that since all the entries of the random matrix are continuous, the probability of the eigenvalues to be equal is 0.

(a) Since this is a 2×2 matrix an explicit calculation of the characteristic polynomial gives

$$p(t) = t^2 - (X_1 + X_2)t + X_1X_2 - X_3^2.$$

The roots are

$$\lambda_1 = \frac{X_1 + X_2 + \sqrt{(X_1 + X_2)^2 - 4X_1X_2 + 4X_3^2}}{2} \text{ and } \lambda_1 = \frac{X_1 + X_2 - \sqrt{(X_1 + X_2)^2 - 4X_1X_2 + 4X_3^2}}{2}.$$

This gives $s = \sqrt{(X_1 - X_2)^2 + 4X_3^2}$.

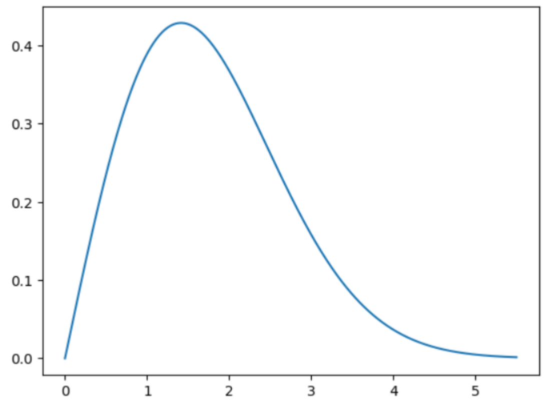

Figure 1: Wigner's surmise

(b) Notice that $X_1 - X_2 \sim_d \mathcal{N}(0,2)$ and $X_3 \sim_d \mathcal{N}(0,2)$. Hence $s \sim_d \sqrt{2}\sqrt{U^2+V^2}$ where U and V are independent standar normals. We get that $s \sim_d \sqrt{2}\chi(2)$, so s has the same distribution as a reescaled by $\sqrt{2}$ chi-2 distribution. The pdf of a chi-2 distribution is given by $p(x) = xe^{-x^2/2} I_{x\geq 0}$. We conclude that the density function of s is given by $p(s) = \frac{s}{2} e^{s^2/4} I_{s\geq 0}$.

(c) Using python we obtain a nice graph. The observation is that while the probability of the eigenvalues being far away decreases exponentially, the probability of them being arbitrarily close goes to 0.

6. **(1D Gaussian process)** In this problem, you will implement a 1D Gaussian process that predicts outputs based on noisy training data. You will be given (noisy) 1D training data pairs $D_{\text{train}} = \{(x_1, y_1), (x_2, y_2) \ldots\}$. Your task is to predict the output for a set of test queries $D_{\text{test}} = \{x_1^*, x_2^*, \ldots\}$, conditioned on the training data. Implement two separate kernel functions, namely the

- **Squared Exponential Kernel**: This is the kernel we discussed in class.

$$k(x_i, x_j) = \sigma_f^2 \exp\left(-\frac{(x_i - x_j)^T M (x_i - x_j)}{2}\right)$$

where σ_f is a scale factor for the kernel and M is a metric measuring distance between two input vectors. In the 1D case, $M = \frac{1}{l^2}$ where l is the length scale of the kernel.

- **Matérn Kernel**: This kernel is used commonly in many machine learning applications.

$$k(x_i, x_j) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} r}{l}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu} r}{l}\right)$$

where ν and l are (positive) parameters of the kernel and $r = |x_i - x_j|$. K_ν is a modified bessel function and Γ is the gamma function. Good parameters settings for ν are 0.25 - 3. You can use scipy.special.kv() in Python or besselK() in R for implementing K_ν.

(a) Implement the squared exponential and Matérn kernel functions to compute similarity between any pair of inputs. The output for each function should be a kernel matrix K.

(b) Using your kernel functions, implement a Gaussion process regression function to predict the posterior mean and variance of test data $\vec{y}^*$.

(c) The simulation function and plotting function are provided in the file `ps5_GP_1D.ipynb`. Vary the kernel parameters (e.g., σ_f, l, and ν) and observe how they affect the predictive mean and variance. What impact do these parameters have on the smoothness and uncertainty of your GP predictions?

Note: It's recommended to use Python (Jupyter notebook) and submit a pdf file including code, plots and comments. If you prefer using another coding language, please make sure the data simulation is the same with the provided code.