STAT 542: Homework 6

Ahmadreza Eslaminia (ae15)

Please make sure that your solutions are readable and the file size is reasonable. Typing the answers is highly encouraged.

Problem 1.

The following example, due to Fellen, shows that different probability distributions can have exactly the same finite integer moments: Let $Z = e^{-F}$, where $F \sim \mathcal{N}(0, 1)$ is standard normal.

- [$1pts$] Compute all the integer moments of Z.
- [2 pts] Consider the parameter family with density given by $p_\theta(F) = p(F)[1 + \theta \sin(2\pi F)]$, where $|\theta| < 1$ and $p()$ denotes the density of the standard normal distribution. Compute all the integer moments of Z_θ.
- [2 pts bonus] Is it possible to estimate θ using the empirical moments of samples from $Z_\theta = e^{-F_\theta}$ where $F_\theta \sim p_\theta$? Why? Hint: Do the integer moments of Z_θ depend on θ ? Is Carleman's condition (p10 of the slides) satisfied?

Solution 1

part 1

To calculate the nth moment of $Z = e^{-F}$ where $F \sim \mathcal{N}(0, 1)$, we need to evaluate the expected value $E[Z^n]$:

$$\begin{aligned} E[Z^n] &= \int_{-\infty}^{\infty} (e^{-F})^n \frac{1}{\sqrt{2\pi}} e^{-\frac{F^2}{2}} dF \\ &= \int_{-\infty}^{\infty} e^{-nF} \frac{1}{\sqrt{2\pi}} e^{-\frac{F^2}{2}} dF \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{F^2}{2} - nF} dF. \end{aligned}$$

Completing the square in the exponent gives us $-\frac{(F+n)^2}{2} + \frac{n^2}{2}$. Thus, the integral becomes

$$
\begin{aligned}
E[Z^n] &= \frac{1}{\sqrt{2\pi}} e^{\frac{n^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(F+n)^2}{2}} dF \\
&= e^{\frac{n^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du \quad \text{where } u = F + n \\
&= e^{\frac{n^2}{2}} \quad \text{(since the integral of the PDF of a standard normal over all } \mathbb{R} \text{ is 1).}
\end{aligned}
$$

Therefore, the nth moment of Z is $E[Z^n] = e^{\frac{n^2}{2}}$ for any integer n.

part 2

Given a parameter family with density $p_\theta(F) = p(F)[1 + \theta \sin(2\pi F)]$, where $|\theta| < 1$ and $p(\cdot)$ denotes the density of the stanard normal distribution, we compute all the integer moments of $Z_\theta = e^{-F}$.

The k^{th} moment of Z_θ is given by:

$$
\begin{aligned}
E[Z_\theta^k] &= E\left[e^{-kF}\right] \\
&= \int_{-\infty}^{\infty} e^{-kF} p_\theta(F)\, dF
\end{aligned}
$$

Substituting the given density $p_\theta(F)$, we get:

$$
\begin{aligned}
E[Z_\theta^k] &= \int_{-\infty}^{\infty} e^{-kF} \left[p(F) + \theta p(F) \sin(2\pi F)\right] \, dF \\
&= E\left[e^{-kF}\right] + \theta \int_{-\infty}^{\infty} e^{-kF} p(F) \sin(2\pi F)\, dF \\
&= e^{k^2/2} + \theta I_k
\end{aligned}
$$

where

$$
I_k = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-kF - \frac{F^2}{2}} \sin(2\pi F)\, dF.
$$

Using the property of the standard normal distribution:

$$
\int_{-\infty}^{\infty} x^n e^{-x^2/2}\, dx = (n-1)!! \sqrt{2\pi} e^{-n^2/4}
$$

and letting $x = F$ and $n = 2k + 1$ (since $\sin(2\pi F)$ is an odd funcion and we integrate over the whole real line, which contributes to even powers of F vanish), we get:

$$
\begin{aligned}
I_k &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-kF} e^{-F^2/2} \sin(2\pi F)\, dF \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F^k e^{-F^2/2} \sin(2\pi F)\, dF \\
&= \frac{1}{\sqrt{2\pi}} (k-1)!! \sqrt{2\pi} e^{-(k)^2/4} \quad \text{(for odd } k \text{ using the provided standard normal distribution property).}
\end{aligned}
$$

Therefore, for k , the k^{th} integer moment of Z_θ is given by:

$$E[Z_\theta^k] = e^{k^2/2} + \theta(k-1)!! e^{-k^2/4},$$

The last step simplification assumes the integral converges, which is not generally the case due to the oscillatory nature of the sine function.

part 3

Given the k^{th} integer moment of Z_θ as:

$$E[Z_\theta^k] = e^{k^2/2} + \theta(k-1)!! e^{-k^2/4},$$

we consider whether it is possible to estimate θ using the empirical moments from samples of Z_θ.

To estimate θ from the empirical moments, the moments must uniquely determine the distribution of Z_θ. Carleman's condition provides a sufficient criterion for the moments to uniquely determine the distribution. It states that if the series

$$\sum_{r=1}^{\infty} M_r^{-\frac{1}{2r}}$$

diverges, where M_r are the moments of the distribution, then the moments uniquely determine the distribution.

For the moments given, we need to assess whether the growth rate of the moments satisfies Carleman's cndition. The presence of the exponential term $e^{-k^2/4}$ in the expression for the moments raises a concern that the moments may grow too quickly to satisfy Carleman's condition.

To conclude whether θ can be estimated from the empirical moments, we must evaluate the series given by Carleman's condition with the explicit moments obtained in part 2. If the series diverges, the moments uniquely determine the distribution, and θ can theoretically be estimated from the empirical moments. If the series converges, or if the growth rate of the moments is too fast, the moments may not uniquely determine the distribution of Z_θ, complicating the estimation of θ from the empirical moments alone.

Problem 2.

[2pts] Consider a naive language generation model that produces a random text composed only of the three words 'machines', 'have', and 'conscious'. Using the bag of words representation, we may assume that the words are generated indepndently and identically distributed (i.i.d.) according to some unknown distribution $[p_1, p_2, p_3]$ over those three words.

Without any further information, a statistician assumes that the prior distribution of $[p_1, p_2, p_3]$ is uniform on the probability simplex (i.e., the triangle formed by the vertices $[1, 0, 0], [0, 1, 0]$, and $[0, 0, 1]$). Then the statistician obtains a text in which the number of the three words 'machines', 'have', and 'conscious' are 25, 25, 50. What is the posterior distribution of $[p_1, p_2, p_3]$?

Hint: Use the fact that the Dirichlet distribution is the Bayesian conjugate of the multinomial distribution.

Solution 2

Given no prior information, we assume that the prior distribution of $[p_1, p_2, p_3]$ is uniform on the probability simplex. This is equivalent to a Dirichlet distribution with parameters $\alpha_1 = \alpha_2 = \alpha_3 = 1$.

After observing a text with counts of 'machines', 'have', and 'conscious' being 25, 25, and 50 respectively, we update our prior to get the posterior distribution. The Dirichlet posterior parameters are updated by adding the observed counts to the prior parameters:

$$\begin{aligned}
\alpha_1' &= \alpha_1 + \text{count of 'machines'} = 1 + 25, \\
\alpha_2' &= \alpha_2 + \text{count of 'have'} = 1 + 25, \\
\alpha_3' &= \alpha_3 + \text{count of 'conscious'} = 1 + 50.
\end{aligned}$$

Hence, the updated parameters are $\alpha' = [26, 26, 51]$. The posterior distribution for $[p_1, p_2, p_3]$ is then a Dirichlet distribution with these parameters:

$$\text{Dirichlet}(\alpha') = \text{Dirichlet}(26, 26, 51).$$

The expected values of the probabilities $[p_1, p_2, p_3]$ under this posterior distribution are calculated as:

$$\begin{aligned}
E[p_1] &= \frac{\alpha_1'}{\sum_{k=1}^{3} \alpha_k'} = \frac{26}{103}, \\
E[p_2] &= \frac{26}{103}, \\
E[p_3] &= \frac{51}{103}.
\end{aligned}$$

This provides the Bayesian update of the distribution of $[p_1, p_2, p_3]$ after observing the text.

[1] Feller, 1968, An Introduction to Probability Theory and Its Applications.

Problem 3.

The following example shows that the nonnegative rank of bounded-rank matrices may be unbounded: Let $M \in \mathbb{R}^{n \times n}$ where $M_{ij} = (i-j)^2$.

1pts Show that $\operatorname{rank}(M) \leq 3$ Hint: it suffices to show that $M_{ij} = \sum_{k=1}^{3} f_k(i) g_k(j)$ for some functions f_k and $g_k, k = 1, 2, 3$.

1pts Show that $\operatorname{rank}(M) \geq 3$ for $n \geq 3$. Hint: find $a \times 3$ submatrix of M which has rank 3.

2pts bonus Show that the nonnegative rank [2] $\operatorname{rank}^+(M) \geq \log_2 n$ Hint: It suffices to consider just the locations of zeros. Suppose that $M = AB^\top$ for some entrywise nonnegative matrices $A, B \in \mathbb{R}^{n \times r}$. Can two rows of A have the same zero pattern?

Remark: This problem illustrates that for general nonnegative M, the nonnegative rank can be much larger than the rank. Curiously, if M is restricted to have 0 or 1 entries, the problem becomes much harder, and it is not known whether $\log \operatorname{rank}^+(M) < (\log \operatorname{rank}(M))^{O(1)}$! This is a reformulation of the famous log-rank conjecture; see Moitra's book.

Solution 3

part 1

Given that $M_{ij} = (i-j)^2$, we can express M_{ij} as:

$$M_{ij} = i^2 - 2ij + j^2.$$

To express M_{ij} as a sum of products of functions of i and j, we define:

$$\begin{aligned} f_1(i) &= i^2, \quad g_1(j) = 1, \\ f_2(i) &= i, \quad g_2(j) = -2j, \\ f_3(i) &= 1, \quad g_3(j) = j^2. \end{aligned}$$

Then M_{ij} can be written as:

$$M_{ij} = \sum_{k=1}^{3} f_k(i) g_k(j).$$

This means we can construct matrix M as the product of two matrices F and G where:

$$F_{ik} = f_k(i) \quad \text{and} \quad G_{kj} = g_k(j),$$

with $k = 1, 2, 3$.

Thus, matrix M can be written as:

$$M = F \cdot G.$$

Since matrices F and G have at most 3 columns and 3 rows respectively, the product $F \cdot G$ will have a rank that is less than or equal to 3. This proves that $\text{rank}(M) \leq 3$.

part 2

We have

$$M_3 = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix}$$

We calculate the entries of M_3 as follows:

$$M_3 = \begin{pmatrix} (1-1)^2 & (1-2)^2 & (1-3)^2 \\ (2-1)^2 & (2-2)^2 & (2-3)^2 \\ (3-1)^2 & (3-2)^2 & (3-3)^2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 4 \\ 1 & 0 & 1 \\ 4 & 1 & 0 \end{pmatrix}$$

The determinant of M_3 is computed as:

$$\det(M_3) = \begin{vmatrix} 0 & 1 & 4 \\ 1 & 0 & 1 \\ 4 & 1 & 0 \end{vmatrix} = (0)(0)(0)+(1)(1)(4)+(4)(1)(1)-(4)(0)(0)-(1)(1)(0)-(0)(1)(4) = 8$$

Since $\det(M_3) = 8 \neq 0$, the submatrix M_3 has full rank, i.e., $\text{rank}(M_3) = 3$. Therefore, the rank of the full matrix M must be at least 3 for $n \geq 3$. This concludes the proof that $\text{rank}(M) \geq 3$.

part 3

Assume that two rows of A, say a_i^T and a_j^T, have the same zero pattern. Let S be the set of indices of non-zero entries in a_i^T (and a_j^T). Then, we can write:

$$a_i^T = (a_{i1}, a_{i2}, \ldots, a_{in}) = (x_1, x_2, \ldots, x_k, 0, \ldots, 0)$$
$$a_j^T = (a_{j1}, a_{j2}, \ldots, a_{jn}) = (y_1, y_2, \ldots, y_k, 0, \ldots, 0)$$

where $k = |S|$, and $x_l, y_l > 0$ for $l = 1, 2, \ldots, k$.

Consider the i-th row of M, denoted by m_i^T. We have:

$$m_i^T = a_i^T B = (x_1, x_2, \ldots, x_k, 0, \ldots, 0)B$$

Similarly, for the j-th row of M, we have:

$$m_j^T = a_j^T B = (y_1, y_2, \ldots, y_k, 0, \ldots, 0)B$$

Since $x_l, y_l > 0$ for $l = 1, 2, \ldots, k$, and B is entrywise nonnegative, we can conclude that $m_i^T = m_j^T$.

This means that if two rows of A have the same zero pattern, then the corresponding rows of $M = AB^T$ will be equal. However, this contradicts the assumption that M has rank at least $\log_2 n$.

Therefore, for M to have rank at least $\log_2 n$, no two rows of A can have the same zero pattern. This implies that the number of distinct zero patterns in the rows of A must be at least $\log_2 n + 1$.

Since each row of A has n entries (either zero or non-zero), the total number of possible zero patterns is 2^n. Hence, the number of distinct zero patterns in the rows of A must satisfy $\log_2 n + 1 \leq 2^n$. This implies that rank $^2\,\mathrm{rank}^+(M) \geq \log_2 n$, as required.