

Peyman Naseri, Matin Ale Mohammad, Sana Harighi, Alireza Gargoori, Maryam Hokmabadi

1 Probability & Statistics

1.1

Suppose X_1, X_2, \dots are independent normal random variables with mean 0 and variance 9. N is also an integer random variable, independent of X_i s, with mean 2 and variance 1. We define $S \triangleq \sum_{i=1}^N X_i$.

- Prove: $Var(X) = \mathbb{E}(Var(X|Y)) + Var(\mathbb{E}(X|Y))$

Solution :

$$\begin{aligned} Var(X|Y) &= E[X^2|Y] - (E[X|Y])^2 \\ \rightarrow E[Var(X|Y)] &= E[E[X^2|Y]] - E[(E[X|Y])^2] = E[X^2] - E[(E[X|Y])^2] \\ Var(E[X|Y]) &= E[(E[X|Y])^2] - (E[E[X|Y]])^2 = E[(E[X|Y])^2] - (E[X])^2 \\ \rightarrow E[Var(X|Y)] + Var(E[X|Y]) &= E[X^2] - (E[X])^2 \end{aligned}$$

- Find the variance of S .

Solution :

$$\begin{aligned} Var(S) &= E(S^2) - E^2(S) (*) \\ E(A) &= E_B(E_A[A|B])(**) \\ \xrightarrow{(*),(**)} Var(S) &= E_N[E_S^2[S^2|N]] - (E_N[E_S[S|N]])^2 \end{aligned}$$

$$\begin{aligned} E[S|N] &= E\left[\sum_{i=1}^N X_i\right] = \sum_{i=1}^N E[X_i] = 0(1) \\ E[S^2|N] &= E\left[\sum_{i=1}^N \sum_{j=1}^N X_i X_j\right] \\ &= \sum_{i=1}^N \sum_{j=1}^N E[X_i X_j] \\ &= \sum_{i=1}^N E[X_i^2] = N * Var(X_i) = 9 * N(2) \\ (1), (2) \rightarrow Var(S) &= E[9 * N] + E[0] \\ &= 9 * E[N] = 9 * 2 = 18 \end{aligned}$$

- Derive the correlation coefficient between the random variables S and N .

Solution :

To derive the correlation coefficient between the random variables S and N , we need to first find their covariance and then divide it by the product of their standard deviations. Using the linearity of covariance, we have:

$$\text{cov}(S, N) = \text{cov}\left(\sum_{i=1}^N X_i, N\right) = \sum_{i=1}^N \text{cov}(X_i, N)$$

Since X_i and N are independent, their covariance is zero. Thus, we have:

$$\text{cov}(S, N) = \sum_{i=1}^N \text{cov}(X_i, N) = \sum_{i=1}^N 0 = 0$$

Next, we need to find the standard deviations of S and N . Since the X_i are independent and identically distributed, we have:

$$\text{Var}(S) = \sum_{i=1}^N \text{Var}(X_i) = \sum_{i=1}^N 9 = 9N$$

Taking the square root of both sides, we get:

$$\text{SD}(S) = 3\sqrt{N}$$

Similarly, we have:

$$\text{SD}(N) = \sqrt{\text{Var}(N)} = \sqrt{1} = 1$$

Therefore, the correlation coefficient between S and N is:

$$\rho_{S,N} = \frac{\text{cov}(S, N)}{\text{SD}(S) \text{SD}(N)} = \frac{0}{3\sqrt{N} \cdot 1} = 0$$

1.2

Suppose the data samples of an experiment are drawn from a normal distribution $\mathcal{N}(\mu, \sigma^2)$. We assume that the variance of this population σ^2 , is known, but its mean μ , is unknown, and we want to estimate it from N independent observations X_1, \dots, X_N .

- Find the ML estimate for the population mean.

Solution :

The likelihood function is:

$$L(\mu, \sigma^2; x_1, x_2, \dots, x_n) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2\right)$$

And the log likelihood is:

$$\ln L(\mu, \sigma^2; x_1, x_2, \dots, x_n) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2$$

$$\begin{aligned} \hat{\mu}_{ML} = \operatorname{argmax}_{\mu} \ln L(\mu, \sigma^2; x_1, x_2, \dots, x_n) &\Rightarrow \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2; x_1, x_2, \dots, x_n) = 0 \\ &\Rightarrow 2 \sum_{i=1}^n \frac{(x_i - \hat{\mu}_{ML})}{2\sigma^2} = 0 \end{aligned}$$

$$\sum_{i=1}^n (x_i - \hat{\mu}_{ML}) = 0 \Rightarrow \sum_{i=1}^n x_i - n\hat{\mu}_{ML} = 0$$

So the MLE is calculated as below:

$$\hat{\mu}_{MLE} = \frac{1}{n} \sum_{i=1}^n x_i : \text{Sample mean}$$

- Suppose that the prior distribution of the parameter μ is the normal distribution $\mathcal{N}(\alpha, \beta^2)$. Find the MAP estimate for the population mean. What is the effect of choosing this prior on the posterior distribution?

Solution :

$$\mu_{\text{prior}} \sim \mathcal{N}(\alpha, \beta^2) \Rightarrow f_U(\mu) = \frac{1}{2\pi\beta} \exp\left(-\frac{(\mu - \alpha)^2}{2\beta^2}\right)$$

$$\begin{aligned} \hat{\mu}_{MAP} &= \operatorname{argmax}_{\mu} L(\mu, \sigma^2; x_1, x_2, \dots, x_n) f_U(\mu) \\ \hat{\mu}_{MAP} = \operatorname{argmax}_{\mu} \ln(L(\mu, \sigma^2; x_1, x_2, \dots, x_n) f_U(\mu)) &= \operatorname{argmax}_{\mu} [\ln(L(\mu, \sigma^2; x_1, x_2, \dots, x_n)) + \ln(f_U(\mu))] \end{aligned}$$

$$\ln(L(\mu, \sigma^2; x_1, x_2, \dots, x_n) f_U(\mu)) = -\frac{N}{2} \ln(2\pi) - N \ln(\sigma) - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}$$

$$-\frac{1}{2} \ln(2\pi) - \ln(\beta) - \left(\frac{(\mu - \alpha)^2}{2\beta^2}\right) = H(\mu, \sigma^2; x_1, x_2, \dots, x_n)$$

$$\Rightarrow \frac{\partial}{\partial \mu} H(\mu, \sigma^2; x_1, x_2, \dots, x_n) = 0$$

$$\Rightarrow \frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \hat{\mu}_{MAP}) - \frac{1}{\beta^2} (\hat{\mu}_{MAP} - \alpha) = 0$$

$$\Rightarrow \hat{\mu}_{MAP} = \frac{\frac{1}{\sigma^2} \sum_{i=1}^n x_i + \frac{\alpha}{\beta^2}}{\frac{\sigma^2 + N\beta^2}{(\sigma\beta)^2}} = \frac{\beta^2 \sum_{i=1}^n x_i + \alpha\sigma^2}{\sigma^2 + n\beta^2}$$

- Explain the relationship between these two estimates when the number of samples is increased.

Solution :

$$\begin{aligned}
\hat{\mu}_{MLE} &= \frac{1}{n} \sum_{i=1}^n x_i : \text{Sample mean} \\
\hat{\mu}_{MAP} &= \frac{\beta^2 \sum_{i=1}^n x_i + \alpha \sigma^2}{\sigma^2 + n\beta^2} \\
\lim_{n \rightarrow \infty} \hat{\mu}_{MAP} &= \frac{\beta^2 \sum_{i=1}^n x_i}{N\beta^2} = \frac{1}{n} \sum_{i=1}^n x_i = \text{Sample Mean} \\
&\Rightarrow \lim_{n \rightarrow \infty} \hat{\mu}_{MAP} = \hat{\mu}_{ML}
\end{aligned}$$

It can be shown that generally this statement is true, by increasing samples MAP and ML estimators will approach towards each other.

1.3

Suppose X_1, X_2, \dots, X_n are iid random variables. Find the probability density function of $Y_1 = \max[X_1, X_2, \dots, X_n]$ and $Y_2 = \min[X_1, X_2, \dots, X_n]$.

Solution :

$$\begin{aligned}
Y_1 &= \max[X_1, X_2, X_3, \dots, X_n] \\
CDF_{of} Y_1 : F_{Y_1}(y_1) &= P(Y_1 \leq y_1) = P(\max[X_1, X_2, X_3, \dots, X_n] \leq y_1) \\
\Rightarrow F_{Y_1}(y_1) &= P(X_1 \leq y_1, X_2 \leq y_1, \dots, X_n \leq y_1) \\
&= P(X_1 \leq y_1) * P(X_2 \leq y_1) * \dots * P(X_n \leq y_1) \\
&= F_{X_1}(y_1) F_{X_2}(y_1) \dots F_{X_n}(y_1) = F_X^n(y_1) \\
PDF_{of} Y_1 : \frac{\partial F_{Y_1}(y_1)}{\partial y_1} &= \frac{\partial F_X^n(y_1)}{\partial y_1} \\
&= n F_X^{(n-1)}(y_1) f_X(y_1)
\end{aligned}$$

$$\begin{aligned}
Y_2 &= \min[X_1, X_2, X_3, \dots, X_n] \\
CDF_{of} Y_2 : F_{Y_2}(y_2) &= P(Y_2 \leq y_2) = P(\min[X_1, X_2, X_3, \dots, X_n] \leq y_2) \\
&= 1 - P(\min[X_1, X_2, X_3, \dots, X_n] > y_2) \\
&= 1 - P(X_1 > y_2, X_2 > y_2, \dots, X_n > y_2) \\
&= 1 - P(X_1 > y_2) P(X_2 > y_2) \dots P(X_n > y_2) \\
&= 1 - (1 - P(X_1 \leq y_2))(1 - P(X_2 \leq y_2)) \dots (1 - P(X_n \leq y_2)) \\
&= 1 - (1 - F_{X_1}(y_2))(1 - F_{X_2}(y_2)) \dots (1 - F_{X_n}(y_2)) \\
&= 1 - (1 - F_X(y_2))^n \\
PDF_{of} Y_2 : \frac{\partial F_{Y_2}(y_2)}{\partial y_2} &= \frac{\partial [1 - (1 - F_X(y_2))^n]}{\partial y_2} \\
&= -n(1 - F_X(y_2))^{n-1} f_X(y_2)
\end{aligned}$$

2 Linear Algebra

2.1

Suppose $A, B \in M_{m \times n}(\mathbb{R})$. Show that $\langle A, B \rangle = \text{tr}(B^T A)$ is an inner product.

Solution :

First we prove that $\langle A, A \rangle \geq 0$ and $\langle A, A \rangle = 0$; iff $A = 0$:

$$\langle A, A \rangle = \text{tr}(A^T A) = \sum_{i=1}^n (A^T A)_{i,i} = \sum_{i=1}^n \left(\sum_{j=1}^m A_{i,j}^T * A_{j,i} \right) = \sum_{i=1}^n \left(\sum_{j=1}^m A_{j,i} * A_{j,i} \right) = \sum_{j=1}^m \sum_{i=1}^n A_{j,i}^2$$

Due to the fact that the above expression is a summation of some non-negative value, it will always be non-negative and zero, if and only if all of its values are zero, which means $A = 0$.

Now we prove this operation has the property of addition to the first component:

$$\langle A_1 + A_2, B \rangle = \text{tr}(B^T (A_1 + A_2)) = \text{tr}(B^T A_1 + B^T A_2) = \text{tr}(B^T A_1) + \text{tr}(B^T A_2) = \langle A_1, B \rangle + \langle A_2, B \rangle$$

Then Homogeneity to the first component needs to be proved as below:

$$\langle \lambda A, B \rangle = \text{tr}(B^T (\lambda A)) = \text{tr}(\lambda (B^T A)) = \lambda \text{tr}(B^T A) = \lambda \langle A, B \rangle$$

Now check conjugate symmetry. since we are in Real numbers, we have to prove $\langle A, B \rangle = \langle B, A \rangle$:

$$\langle A, B \rangle = \text{tr}(B^T A) = \sum_{i=1}^n (B^T A)_{i,i} = \sum_{i=1}^n \left(\sum_{j=1}^m B_{i,j}^T * A_{j,i} \right) = \sum_{i=1}^n \left(\sum_{j=1}^m B_{j,i} * A_{j,i} \right) = \sum_{i=1}^n (A^T B)_{i,i} = \text{tr}(A^T B) = \langle B, A \rangle$$

Now the statement is proved.

2.2

Assume that x is a vector and A is a square matrix. Show that:

- $\frac{\partial x^T A x}{\partial x} = 2x^T A$

Solution :

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

So, we will have:

$$x^T A x = \begin{bmatrix} \sum_{i=1}^n a_{i1} x_i & \sum_{i=1}^n a_{i2} x_i & \dots & \sum_{i=1}^n a_{in} x_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$$

Therefore,

$$\frac{\partial x^T A x}{\partial x_i} = 2x^T A = \sum_{j=1}^n (a_{ij} + a_{ji})x_j \frac{\partial x^T A x}{\partial x} = x^T (A + A^T) = 2x^T A$$

Notice that A is symmetric

- $\frac{\partial \text{trace}(x^T A x)}{\partial x} = x^T (A + A^T)$

Solution :

We have:

$$\frac{\partial}{\partial x} \text{trace}(x^T A x) = \frac{\partial}{\partial x} \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j,$$

where n is the dimension of x and A .

Since the trace is just the sum of the diagonal entries of a matrix, we can rewrite this as:

$$\frac{\partial}{\partial x} \text{trace}(x^T A x) = \frac{\partial}{\partial x} \text{trace}(x x^T A)$$

Using the cyclic property of the trace, we can move A to the front:

$$\frac{\partial}{\partial x} \text{trace}(x^T A x) = \frac{\partial}{\partial x} \text{trace}(A x x^T)$$

Using the formula for the derivative of the trace, we have:

$$\frac{\partial}{\partial x} \text{trace}(x^T A x) = A^T x + A x$$

Note that $A^T = A$ since A is a square matrix. Multiplying x^T on the left, we get:

$$\frac{\partial}{\partial x} \text{trace}(x^T A x) = x^T A^T + x^T A = x^T (A + A^T)$$

Therefore, we have shown that:

$$\frac{\partial}{\partial x} \text{trace}(x^T A x) = x^T (A + A^T)$$

2.3

Suppose $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of matrix A , Prove:

- $\lambda_1 + \lambda_2 + \dots + \lambda_n = \text{trace}(A)$

Solution :

Using the definition of the trace of a matrix, we have:

$$\text{trace}(A) = a_{11} + a_{22} + \dots + a_{nn},$$

where a_{ii} is the (i, i) -th entry of A .

Every matrix satisfies its own characteristic polynomial, which is defined as $\det(A - \lambda I) = 0$, where I is the identity matrix and λ is an eigenvalue of A . So, we can write:

$$\det(A - \lambda I) = (-1)^n (\lambda^n + c_{n-1} \lambda^{n-1} + \dots + c_1 \lambda + c_0),$$

where n is the order of A and c_i are constants. Substituting $\lambda = 0$, we get:

$$\det(A) = (-1)^n c_0,$$

which implies that c_0 is equal to $(-1)^n$ times the determinant of A . Substituting $\lambda = a_{ii}$, we get:

$$\det(A - a_{ii}I) = (-1)^n (a_{ii} - a_{11})(a_{ii} - a_{22}) \dots (a_{ii} - a_{nn}),$$

which implies that the product of the eigenvalues of A is equal to $(-1)^n$ times the determinant of A . Dividing both sides by $(-1)^n$, we get:

$$\lambda_1 \lambda_2 \dots \lambda_n = \det(A),$$

where $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of A . Using the above result, we can express the characteristic polynomial of A as:

$$\det(A - \lambda I) = (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n),$$

and we can expand this polynomial as:

$$\det(A - \lambda I) = \lambda^n + (-1)^{n-1} (\lambda_1 + \lambda_2 + \dots + \lambda_n) \lambda^{n-1} + \dots + (-1)^n \det(A).$$

Comparing the coefficients of λ^{n-1} on both sides, we get:

$$-(\lambda_1 + \lambda_2 + \dots + \lambda_n) = \text{trace}(A),$$

which implies that the sum of the eigenvalues of A is equal to the trace of A . Therefore, we have proven the desired result.

- $\lambda_1 \lambda_2 \dots \lambda_n = \det(A)$

Solution :

Since λ s are eigenvalues of matrix, then they are also the roots of characteristic polynomial.

$$\det(A - \lambda I) = p(\lambda) = (-1)^n (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda),$$

We put λ equal to zero. So:

$$\det(A) = \lambda_1 \lambda_2 \dots \lambda_n,$$

- AB and BA have the same set of eigenvalues.

Solution :

To prove that AB and BA have the same eigenvalues, we need to show that the characteristic polynomials of them are equal. The characteristic polynomial of a matrix is defined as: $\det(\lambda I - A)$

where λ is an eigenvalue of A and I is the identity matrix of the same size as A . So, we need to show that for all λ :

$$\det(\lambda I - AB) = \det(\lambda I - BA)$$

Let λ be an arbitrary eigenvalue of AB , with corresponding eigenvector \mathbf{v} . Then we have:

$$AB\mathbf{v} = \lambda\mathbf{v} \tag{1}$$

Let $\mathbf{u} = B\mathbf{v}$. Multiplying both sides by B , we get:

$$BA\mathbf{u} = BAB\mathbf{v} = B(AB\mathbf{v}) = B(\lambda\mathbf{v})$$

Since B commutes with λI , we can write this as:

$$BA\mathbf{u} = \lambda(B\mathbf{v}) = \lambda\mathbf{u}$$

Therefore, we have:

$$BA\mathbf{u} = \lambda\mathbf{u}$$

This shows that λ is an eigenvalue of BA with corresponding eigenvector $\mathbf{u} = B\mathbf{v}$.

Therefore, every eigenvalue of AB is also an eigenvalue of BA , and vice versa. Since the characteristic polynomial of a matrix is the product of its eigenvalues (with appropriate algebraic multiplicities), it follows that the two matrices have the same characteristic polynomial, and hence the same eigenvalues.

- A and A^T have the same set of eigenvalues.

Solution :

We know the eigenvalues are the roots of the characteristic polynomial; therefore, if we prove A and A^T have the same characteristic polynomial equation, the statement will be proved.

$$|A^T - \lambda I| = |A^T - \lambda I^T| = |(A - \lambda I)^T| = |A - \lambda I|,$$