Stats 100B – Intro to Statistics

University of California, Los Angeles

Duc Vu

Summer 2021

This is stats 100B taught by Professor Christou. The formal name of the class is **Introduction to Mathematical Statistics**. There is not an official textbook used for the course. Instead, handouts and reference materials are distributed and can be accessed through the class website. You can find other math/stats lecture notes through my personal blog. Let me know through my email if you notice something mathematically wrong/concerning. Thank you!

Contents

List of Theorems

List of Definitions

§1 Lec 1: Aug 3, 2021

§1.1 Review of Stats 100A

Let X be a random variable.

	Discrete RV	Continuous RV
Distribution Function	pmf	pdf
Expected Value	$EX = \sum_x xp(x)$	$EX = \int_x xf(x)\,dx$
Expectation Function	$Eg(x) = \sum_x g(x)p(x)$	$Eg(x) = \int_x g(x)f(x)dx$
Variance	$EX^2 - (EX)^2$	$EX^2 - (EX)^2$

Let X, Y be random variables with the joint pdf/pmf $f(x, y)$. If X, Y are independent, then

$$f(x, y) = f(x) \cdot f(y)$$

where $f(x)$ is the marginal pdf of x and $f(y)$ is the marginal pdf of y. Also,

$$f(x) = \int_y f(x, y)\,dy$$
$$f(y) = \int_x f(x, y)\,dx$$

Theorem 1.1

X, Y are independent if and only if

$$f(x, y) = g(x) \cdot h(y)$$

Remark 1.2. $g(x)$ and $h(y)$ are not necessarily the marginal pdf of x and y respectively.

Proof. Let $c = \int_x g(x)\,dx$ and $d = \int_y h(y)\,dy$. Notice that

$$c \cdot d = \int_x \int_y \underbrace{g(x)h(y)}_{f(x,y)}\,dx\,dy = 1$$

Now, we find $f(x)$ and $f(y)$

$$f(x) = \int_y f(x, y)\,dy = \int_y g(x)h(y)\,dy = g(x)d$$
$$f(y) = \int_x f(x, y)\,dx = \int_x g(x)h(y)\,dx = h(y)c$$

So,

$$f(x, y) = g(x)h(y)cd = f(x)f(y)$$

Therefore, X, Y are independent. □

Let $X \sim \Gamma(\alpha, \beta)$. Then, for $x > 0, \alpha > 0, \beta > 0$,

$$f(x) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha}$$

where

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$$

We have the following properties

$$\begin{aligned}\Gamma(\alpha+1) &= \alpha\Gamma(\alpha)\\ \Gamma(\alpha+2) &= (\alpha+1)\Gamma(\alpha+1)\\ &= (\alpha+1)\Gamma(\alpha-1)\end{aligned}$$

If α is an integer, then

$$\Gamma(\alpha) = (\alpha-1)!$$

Kernel function of $\Gamma(\alpha, \beta)$ is

$$k(x) = x^{\alpha-1}e^{-\frac{x}{\beta}} = \int_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}\,dx$$

Let's make a substitution $y = \frac{x}{\beta}$. Then,

$$\begin{aligned}\int_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}\,dx &= \int_0^\infty (\beta y)^{\alpha-1}e^{-y}\beta\,dy\\ &= \beta^\alpha\int_0^\infty y^{\alpha-1}e^{-y}\,dy\\ &= \beta^\alpha\Gamma(\alpha)\end{aligned}$$

So

$$\int_0^\infty \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha}\,dx = 1$$

§1.2 Exponential Families

Definition 1.3 (Exponential Family) — A random variable X belongs in the exponential family if its pdf/pmf can be expressed as follows

$$f(x|\theta) = h(x)\cdot c(\theta)\cdot e^{\sum_{i=1}^k w_i(\theta)\cdot t_i(x)}$$

Example 1.4

Let $X \sim b(n,p)$ with n fixed. Show that this belongs in an exponential family.

$$\begin{aligned} p(x) &= \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x}(1-p)^n \left(\frac{p}{1-p}\right)^x \\ &= \binom{n}{x}(1-p)^n e^{\ln\left(\frac{p}{1-p}\right)^x} \\ &= \binom{n}{x}(1-p)^n e^{\left(\ln \frac{p}{1-p}\right)x} \end{aligned}$$

So, we have

$$\begin{aligned} h(x) &= \binom{n}{x} \\ c(\theta) &= (1-p)^n \\ w_1(\theta) &= \ln \frac{p}{1-p} \\ t_1(x) &= x \end{aligned}$$

Notice that in this case we have one parameter, and that is $\theta = p$.

Example 1.5

$X \sim \text{Poisson}(\lambda)$ and

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0,1,2,\ldots$$

Show that it is an exponential family.

$$p(x) = \frac{1}{x!} e^{-\lambda} e^{\ln \lambda^x} = \frac{1}{x!} e^{-\lambda} e^{(\ln \lambda) x}$$

where $h(x) = \frac{1}{x!}$, $c(\theta) = e^{-\lambda}$, $w_1(\theta) = \ln \lambda$, $t_1(x) = x$.

Theorem 1.6 a) $E\left[\sum \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(x)\right] = -\frac{\partial \ln c(\theta)}{\partial \theta_j}$

b) $\text{var}\left(\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(x)\right) = -\frac{\partial^2 \ln c(\theta)}{\partial \theta_j} - E\left[\sum_{i=1}^k \frac{\partial^2 w_i(\theta)}{\partial \theta_j} t_i(x)\right]$

Example 1.7

If $X \sim \text{Poisson}(\lambda)$ then show that $EX = \lambda$. From the theorem above (a)

$$E\left[\frac{1}{\lambda} x\right] = -(-1) \implies EX = \lambda$$

Exercise 1.1. $X \sim N(\mu, \sigma)$. Show that $f(X)$ belongs to an exponential family.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

§1.3 Moment Generating Functions

Definition 1.8 (Moment Generating Function) — Let X be a random variable. Then the mgf of X is

$$M_X(t) = Ee^{tX} = \begin{cases} \int_x e^{tx} f(x)\, dx, & \text{for continuous RV} \\ \sum_x e^{tx} p(x), & \text{for discrete RV} \end{cases}$$

<u>Moments</u>:

$$\begin{aligned} M_X(t) &= \int_x e^{tx} f(x)\, dx \\ M_X'(t) &= \int_x x e^{tx} f(x)\, dx \\ M_X'(0) &= \int_x x f(x)\, dx = EX \\ M_X''(t) &= \int_x x^2 e^{tx} f(x)\, dx \\ M_X''(0) &= \int_x x^2 f(x)\, dx = EX^2 \\ \text{var}(X) &= EX^2 - (EX)^2 \end{aligned}$$

Theorem 1.9
Let $\phi(t) = \ln M_X(t)$. Then

$$\begin{aligned} \phi'(0) &= EX \\ \phi''(0) &= \text{var}(X) \end{aligned}$$

Proof. We have

$$\begin{aligned} \phi'(t) &= \frac{M_X'(t)}{M_X(t)} \\ \phi'(0) &= \frac{M_X'(0)}{M_X(0)} = \frac{E(X)}{1} = EX \end{aligned}$$

and

$$\begin{aligned} \phi''(t) &= \frac{M_X''(t) \cdot M_X(t) - \left(M_X'(t)\right)^2}{\left(M_X(t)\right)^2} \\ &= \dots \\ &= EX^2 - (EX)^2 \\ &= \text{var}(X) \end{aligned}$$

□

The MGF of

- Binomial – $X \sim b(n,p)$

$$
\begin{aligned}
p(x) &= \binom{n}{x} p^x (1-p)^{n-x} \\
M_X(t) = Ee^{tx} &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\
&= (pe^t + 1 - p)^n
\end{aligned}
$$

- Poisson

$$
\begin{aligned}
p(x) &= \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \dots \\
M_X(t) &= \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\
&= e^{-\lambda} e^{\lambda e^t} \\
&= e^{\lambda\left(e^t - 1\right)}
\end{aligned}
$$

- Gamma – $X \sim \Gamma(\alpha, \beta)$, $x, \alpha, \beta > 0$. Note that if $\lambda = 1$ and $\beta = \frac{1}{\lambda}$, then $f(x) = \lambda e^{-\lambda x}$, i.e. exponential distribution.

$$
\begin{aligned}
M_X(t) &= \int_0^{\infty} e^{tx} \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha}\, dx \\
&= \int_0^{\infty} \frac{x^{\alpha-1} e^{-x\left(\frac{1}{\beta} - t\right)}}{\Gamma(\alpha)\beta^\alpha}\, dx
\end{aligned}
$$

Let $y = x\left(\frac{1}{\beta} - t\right)$. Then, after some “massage”, we obtain

$$
M_X(t) = (1 - \beta t)^{-\alpha}
$$

- Exponential – $X \sim \exp(\lambda)$. Then,

$$
M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}
$$

- Normal – $Z \sim N(0,1)$

$$
\begin{aligned}
f(z) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty \\
M_Z(t) = Ee^{tz} &= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz \\
&= e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} \\
&= e^{\frac{1}{2}t^2}
\end{aligned}
$$

Properties of MGF:

Theorem 1.10

If X, Y are independent, then

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

Proof. We have

$$\begin{aligned} M_{X+Y}(t) &= Ee^{t(X+Y)} \\ &= E\left(e^{tX} \cdot e^{tY}\right) \\ &= (Ee^{tX})(Ee^{tY}) \\ &= M_X(t) \cdot M_Y(t) \end{aligned}$$

□

Example 1.11

Let $X_1, X_2, \ldots, X_n$ be i.i.d random variables with $X_i \sim \exp(\lambda)$. Find the distribution of $X_1 + X_2 + \ldots + X_n$. From the theorem above, we have

$$\begin{aligned} M_{X_1+X_2+\ldots+X_n}(t) &= M_{X_1}(t) \cdot M_{X_2}(t) \ldots M_{X_n}(t) \\ &= \left(1 - \frac{t}{\lambda}\right)^{-1} \left(1 - \frac{t}{\lambda}\right)^{-1} \ldots \left(1 - \frac{t}{\lambda}\right)^{-1} \\ &= \left(1 - \frac{t}{\lambda}\right)^{-n} \end{aligned}$$

Thus, the sum $X_1 + X_2 + \ldots + X_n \sim \Gamma\left(n, \frac{1}{\lambda}\right)$.

§2 Lec 2: Aug 4, 2021

§2.1 Moment Generating Functions (Cont'd)

Example 2.1 (Method of MGF)
$X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$, and X, Y are independent.

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= e^{\lambda_1(e^t-1)} \cdot e^{\lambda_2(e^t-1)} \\ &= e^{(\lambda_1+\lambda_2)(e^t-1)} \end{aligned}$$

Thus, $X+Y \sim \text{Poisson}(\lambda_1+\lambda_2)$ (by uniqueness theorem, i.e., each distribution has its own unique generating function).

Example 2.2 (Method of MGF)
Let $X_1, X_2, \ldots, X_n \overset{\text{i.i.d}}{\sim} \text{Poisson}(\lambda)$ and $T = X_1 + X_2 + \ldots + X_n$.

$$\begin{aligned} M_T(t) &= (M_{X_i}(t))^n \\ &= \left(e^{\lambda(e^t-1)}\right)^n \\ &= e^{n\lambda(e^t-1)} \end{aligned}$$

So, $T \sim \text{Poisson}(n\lambda)$.

Example 2.3 (Method of PMF)
From Example 2.1, we have

$$\begin{aligned} P(X+Y=k) &= \sum_{i=0}^{k} p(X=i, Y=k-i) \\ &= \sum_{i=0}^{k} p(X=i) \cdot p(Y=k-i) \\ &= \sum_{i=0}^{k} \frac{\lambda_1^i e^{-\lambda_1}}{i!} \cdot \frac{\lambda_2^{k-i} e^{-\lambda_2}}{(k-i)!} \\ &= e^{-(\lambda_1+\lambda_2)} \sum_{i=0}^{k} \frac{\lambda_1^i \lambda_2^{k-i}}{i!(k-i)!} \cdot \frac{k!}{k!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda_1^i \lambda_2^{k-i} \\ &= \frac{(\lambda_1+\lambda_2)^k e^{-(\lambda_1+\lambda_2)}}{k!} \end{aligned}$$

Thus, $X+Y \sim \text{Poisson}(\lambda_1+\lambda_2)$.

Example 2.4

Suppose $X \sim b(n_1, p)$, $Y \sim b(n_2, p)$, and X, Y are independent. Find the distribution of $X + Y$.

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= (pe^t + 1 - p)^{n_1} \left(pe^t + 1 - p\right)^{n_2} \\ &= \left(pe^t + 1 - p\right)^{n_1 + n_2} \end{aligned}$$

Thus, $X + Y \sim b(n_1 + n_2, p)$.

Properties of MGF:

a) MGF of $X + a$ is

$$\begin{aligned} M_{X+a}(t) &= Ee^{t(X+a} \\ &= e^{ta} \cdot Ee^{tX} = e^{ta} M_X(t) \end{aligned}$$

b) MGF of bX is

$$\begin{aligned} M_{bX}(t) &= Ee^{tbX} \\ &= Ee^{t^* x} \\ &= M_X(t^*) = M_X(bt) \end{aligned}$$

Example 2.5

$X \sim \Gamma(\alpha, \beta)$. Then,

$$M_X(t) = (1 - \beta t)^{-\alpha}$$

Let $Y = cX$ where $c > 0$. We want to find the distribution of Y.

(a) Method of MGF:

$$\begin{aligned} M_Y(t) = M_{cX}(t) &= M_X(ct) \\ &= (1 - c\beta t)^{-\alpha} \end{aligned}$$

Therefore, $Y \sim \Gamma(\alpha, c\beta)$.

(b) Method of CDF:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= p(cX \leq y) \\ &= p(X \leq \frac{y}{c}) \end{aligned}$$

Then, $F_Y(y) = F_X\left(\frac{y}{c}\right)$. Take derivative w.r.t. y

$$\begin{aligned} f_Y(y) &= \frac{1}{c} f_X\left(\frac{y}{c}\right) \\ f(x) &= \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha} \end{aligned}$$

Lastly, replace X with $\frac{Y}{c}$.

c) MGF of $\frac{X+a}{b}$ is

$$\begin{aligned}
M_{\frac{X+a}{b}}(t) &= Ee^{t\cdot\frac{X+a}{b}} \\
&= e^{t\frac{a}{b}} Ee^{t\frac{X}{b}} \\
&= e^{t\frac{a}{b}} \cdot M_X\left(\frac{t}{b}\right)
\end{aligned}$$

Use these properties to find the MGF of $X \sim N(\mu, \sigma)$. Recall that if $Z \sim N(0,1)$, then

$$M_Z(t) = e^{\frac{1}{2}t^2}$$

So, standardizing x to obtain

$$Z = \frac{X-\mu}{\sigma} \implies X = \mu + \sigma Z$$

Then,

$$\begin{aligned}
M_X(t) &= M_{\mu+\sigma Z}(t) \\
&= Ee^{t(\mu+\sigma z)} \\
&= e^{t\mu} M_Z(\sigma t) \\
&= e^{t\mu} e^{\frac{1}{2}t^2\sigma^2}
\end{aligned}$$

Thus, $M_X(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}$.

Example 2.6

Let $X \sim N(\mu_1, \sigma_1)$ and $Y \sim N(\mu_2, \sigma_2)$ and X, Y are independent. We want to find the distribution of $X+Y$.

$$\begin{aligned}
M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\
&= e^{t\mu_1 + \frac{1}{2}t^2\sigma_1^2} \cdot e^{t\mu_2 + \frac{1}{2}t^2\sigma_2^2} \\
&= e^{t(\mu_1+\mu_2) + \frac{1}{2}t^2(\sigma_1^2+\sigma_2^2)}
\end{aligned}$$

Thus, $X + Y \sim N\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$.

Example 2.7

Let $X_1, X_2, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$. Let $T = X_1 + X_2 + \ldots + X_n$. Then

$$\begin{aligned}
M_T(t) &= (M_{X_i}(t))^n \\
&= \left(e^{t\mu + \frac{1}{2}t^2\sigma^2}\right)^n \\
&= e^{tn\mu + \frac{1}{2}t^2 n\sigma^2}
\end{aligned}$$

Thus, $T \sim N(n\mu, \sigma\sqrt{n})$.

Example 2.8

Let $\overline{X} = \frac{\sum X_i}{n} = \frac{T}{n}$. Find $M_{\overline{X}}(t)$.

$$\begin{aligned} M_{\overline{X}}(t) &= M_T\left(\frac{t}{n}\right) \\ &= e^{t\mu + \frac{1}{2}t^2\frac{\sigma^2}{n}} \end{aligned}$$

Therefore, $\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$.

Recall

Theorem 2.9 (Central Limit Theorem)

Let $T = X_1 + \ldots + X_n$ with mean μ and variance σ^2 (can follow any distribution other than normal). As $n \to \infty$,

$$\frac{T - n\mu}{\sigma\sqrt{n}} \to N(0, 1)$$

Proof. Start with the MGF and as $n \to \infty$ we obtain

$$M_{\frac{T-n\mu}{\sigma\sqrt{n}}}(t) \to e^{\frac{1}{2}t^2}$$

□

§2.2 Joint MGF

Let $X = \begin{bmatrix} X_1 & X_2 & \ldots & X_n \end{bmatrix}^\top$ be a random vector and $t = \begin{bmatrix} t_1 & t_2 & \ldots & t_n \end{bmatrix}^\top$.

Definition 2.10 (Joint MGF) — Joint MGF of X is defined as

$$M_X(t) = Ee^{t^\top X} = Ee^{\sum t_i X_i}$$

Let X be a random vector (as above) with mean vector $\mu = \begin{bmatrix} \mu_1 & \mu_2 & \ldots & \mu_n \end{bmatrix}^\top$, i.e.,

$$\mu = EX = E\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} EX_1 \\ EX_2 \\ \vdots \\ EX_n \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

and variance/covariance matrix is defined as

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \ldots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \ldots & \sigma_n^2 \end{bmatrix} = E\left[(X-\mu)(X-\mu)^\top\right]$$

Special Case: For i.i.d random variables,

$$\mu = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} = \mu \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \mu \mathbf{1}$$

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix} = \sigma^2 I$$

Now, let's discuss two results.

1. Let $a = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}^\top$ be a vector of constants. Find the mean and variance of $a^\top X$.

$$\begin{aligned} Ea^\top X &= a^\top EX = a^\top \mu \\ \operatorname{var}(a^\top X) &= E(a^\top X - a^\top \mu)^2 \\ &= a^\top \left[E(X-\mu)(X-\mu)^\top \right] a \\ &= a^\top \Sigma a \end{aligned}$$

or using summation, we have

$$\operatorname{var}(a^\top X) = \sum_{i=1}^{n} a_i^2 \operatorname{var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j>i}^{n} a_i a_j \operatorname{cov}(X_i, X_j)$$

Example 2.11

For $n = 3$,

$$\begin{aligned} \operatorname{var}(a_1X_1 + a_2X_2 + a_3X_3) &= \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \\ &= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + a_3^2\sigma_3^2 + 2a_1a_2\sigma_{12} + 2a_1a_3\sigma_{13} + 2a_2a_3\sigma_{23} \end{aligned}$$

2. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \end{bmatrix}$$

be a $p \times n$ matrix of constants. Find mean and variance of the vector AX.

$$\begin{aligned} E(AX) &= AEX = A\mu \\ \operatorname{var}(AX) &= E\left[(AX - A\mu)(AX - A\mu)^\top \right] \\ &= AE(X-\mu)(X-\mu)^\top A^\top \\ &= A\Sigma A^\top \end{aligned}$$

Consider $X^\top AX$ where $X : n \times 1$, $A : n \times n$ symmetric. For example, $n = 2$,

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}$$

Then $X^\top AX = 5X_1^2 + 3X_2^2 + 4X_1X_2$.

$$\begin{aligned}
E\left[\underbrace{X^\top AX}_{\text{scalar}}\right] &= E\,\mathrm{tr}(X^\top AX) \\
&= E\left(\mathrm{tr}\,AXX^\top\right) \\
&= \mathrm{tr}\left(EAXX^\top\right) \\
&= \mathrm{tr}\left(AEXX^\top\right) \\
&= \mathrm{tr}\left(A(\Sigma + \mu\mu^\top)\right) \\
&= \mathrm{tr}(A\Sigma) + \mathrm{tr}(A\mu\mu^\top) \\
&= \mathrm{tr}(A\Sigma) + \mu^\top A\mu
\end{aligned}$$

Note that $\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB) \neq \mathrm{tr}(BAC)$
Let $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, $t = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$. Then,

$$\begin{aligned}
M_X(t) &= E\left(e^{t_1X_1 + t_2X_2}\right) \\
&= \int_{x_1}\int_{x_2} e^{t_1x_1+t_2x_2} f(x_1,x_2)\,dx_1\,dx_2 \\
M_1(t) = \frac{\partial M_X(t)}{\partial t_1} &= \int_{x_1}\int_{x_2} x_1 e^{t_1x_1+t_2x_2} f(x_1,x_2)\,dx_1\,dx_2
\end{aligned}$$

Set $t = 0$, we obtain

$$\begin{aligned}
M_1(0) &= \int\int x_1 f(x_1,x_2)\,dx_1\,dx_2 \\
&= \int_{x_1} x_1 \left[\int_{x_2} f(x_1,x_2)\,dx_2\right] dx_1 \\
&= \int_{x_1} x_1 f(x_1)\,dx_1 \\
&= EX_1
\end{aligned}$$

So,

$$\begin{aligned}
\mathrm{var}(X_1) &= EX_1^2 - (EX_1)^2 \\
\mathrm{cov}(X_1,X_2) &= E(X_1,X_2) - (EX_1)(EX_2)
\end{aligned}$$