Review Session 4 – Multivariate Gaussians and Random Samples

1 Gaussian Random Vectors

Gaussian random variables and Gaussian random vectors play a central role in statistical theory. This stems from the observation that many noise-like quantities in real world applications appear to behave as Gaussians. Another, perhaps more important reason, is that Gaussian random variables turn out to be remarkably easy to work with and give rise to many elegant, exact results without requiring asymptotics. A third reason why Gaussians are important is that often the estimation/detection problems in the specific Gaussian case greatly inform us about other more general cases. For example, the minimum mean square estimator for Gaussian distributions is the same, and has the same mean square performance, as the linear least squares estimator for other problems with the same mean and covariance. In many cases, first insights into a difficult statistical problem come from understanding the simplified Gaussian problem.

1.1 Preliminaries

The *covariance matrix* (also known as the variance matrix) of a random vector $X \in \mathbb{R}^p$ is a square matrix giving the covariance between each pair of components of X:

$$\mathrm{Cov}(X) \text{ has } (i,j)\text{-th entry } \mathrm{Cov}(X_i, X_j).$$

It is sometimes written as $\mathrm{Var}(X)$ instead since it is really the analogue of the variance of a random variable. Indeed, we may write

$$\mathrm{Cov}(X) = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T] = \mathbb{E}[XX^T] - \mathbb{E}[X] \cdot \mathbb{E}[X]^T.$$

The covariance matrix is a p.s.d. and symmetrix matrix. Furthermore, it behaves similarly to the variance. For a linear transformation of X, $AX + b$ where $A \in \mathbb{R}^{m \times p}$ and $b \in \mathbb{R}^m$,

$$\mathrm{Cov}(AX + b) = A\,\mathrm{Cov}(X)A^T.$$

More generally, the *cross-covariance* of two random vectors $X \in \mathbb{R}^p, Y \in \mathbb{R}^q$ (where p and q are not necessarily the same dimension) is defined as

$$\mathrm{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^T] = \mathbb{E}[XY^T] - \mathbb{E}[X] \cdot \mathbb{E}[Y]^T.$$

Thus, $\mathrm{Cov}(X, Y)$ is a $p \times q$ matrix. It behaves similarly to the covariance of random variables:

1. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)^T$.
2. $\mathrm{Cov}(X, X) = \mathrm{Cov}(X)$.
3. $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$.
4. $\mathrm{Cov}(AX + b, CY + d) = A\,\mathrm{Cov}(X, Y)C^T$ for matrices A, C.
5. If X, Y are independent random vectors, then $\mathrm{Cov}(X, Y) = 0_{p \times q}$.

Now, for a fixed $q \times p$ matrix A, we have that $A \cdot X$ is a linear transformation of X. Then, we have the mean and covariance transform as follows:

1. $\mathbb{E}[AX] = A\mathbb{E}[X]$.
2. $\mathrm{Cov}(AX) = A\,\mathrm{Cov}(X)A^T$.

The *moment generating function* of a random vector $X \in \mathbb{R}^p$ is a multivariate function $M_X : \mathbb{R}^p \to \mathbb{R}$ given by

$$M_X(t) = \mathbb{E}[e^{t^T X}].$$

As in the case of the univariate mgf, we will say the mgf of X exists if $M_X(t)$ is finite in a region around zero, i.e. for all $t \in (-t_0, t_0)^p \subseteq \mathbb{R}^p$. As for random variables, the mgf characterizes the distribution of a random vector.

Now, let's review the spectral decomposition from the first review session on linear algebra. Recall that a symmetric square matrix $\Sigma \in \mathbb{R}^{p\times p}$ has real eigenvalues $\{\lambda_i\}_{i=1}^p$ and a choice of orthogonal eigenvectors $\{u_i\}_{i=1}^p$ (i.e., orthogonal u_i so that $\Sigma u_i = \lambda_i u_i$). Then, letting Λ be the diagonal matrix with (i,i)-th entry λ_i and U be the matrix with columns $u_1, \dots, u_p$, the spectral decomposition gives us

$$\Sigma = \lambda_1 u_1 u_1^T + \cdots + \lambda_p u_p u_p^T = U \Lambda U^T.$$

Furthermore, recall that if Σ is a positive semi-definite or p.s.d. (resp. positive definite) matrix if and only if all the eigenvalues are nonnegative (resp. positive). For p.s.d. Σ, we can define the *nonnegative definite square root* of Σ as $\Sigma^{1/2} := U\Lambda^{1/2}U^T$. This is indeed a square root in the sense that $\Sigma^{1/2} \cdot \Sigma^{1/2} = \Sigma$.

1.2 Gaussian random vectors

First, we define the standard Gaussian vector of dimension p: $Z \sim \mathcal{N}_p(0_p, \mathrm{Id}_p)$ which has mean 0_p and covariance matrix Id_p: this is the random vector $Z = (Z_1, \dots, Z_p)$ such that the Z_i's are i.i.d. $\mathcal{N}(0,1)$ random variables.

Now, given a generic fixed vector $\mu \in \mathbb{R}^p$ and symmetric positive definite square matrix $\Sigma \in \mathbb{R}^{p\times p}$, we say that $X \sim \mathcal{N}_p(\mu, \Sigma)$ if $X = \mu + \Sigma^{1/2} Z$. In this case, we say that X is a *multivariate normal* or *multivariate Gaussian* with mean $\mathbb{E}[X] = \mu$ and covariance matrix $\mathrm{Cov}(X) = \Sigma$. Note that these two facts follow from the definition of X and the linearity of the mean and covariance.

What is the density of the multivariate normal $X \sim \mathcal{N}_p(\mu, \Sigma)$? First, we can determine the density of Z, which is the joint density of p i.i.d. standard normals:

$$f_Z(z) = (2\pi)^{-p/2} \exp\left(-\frac{1}{2}\sum_{i=1}^p z_i^2\right).$$

Then, since $X = \mu + \Sigma^{1/2} Z$, using our multivariate pdf transformation law from the previous review session, we have that the joint density of $X_1, \dots, X_p$ is

$$f_X(x) = (\det(2\pi\Sigma))^{-1/2} \exp\left(-\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)\right).$$

This looks strikingly similar to the univariate Gaussian pdf. The variance again appears in two places: in the normalizing constant and in the exponential now inverted. We call the inverse matrix Σ^{-1} the inverse covariance or the *concentration* matrix.

The standard Gaussian estimation problem is to find μ or Σ. For multivariate Gaussians, μ has p unknown parameters whereas Σ has $\frac{p(p+1)}{2}$ parameters (since Σ is symmetric). Thus, the parameter space might be very large for large p. Because of this, it is customary to reduce the number of parameters by looking at more structured cases:

- Assuming μ lies in a linear subspsace of dimension $r < p$.
- Assuming Σ is *isotropic*: $\Sigma = \sigma^2 \cdot \mathrm{Id}_{p\times p}$.
- Assuming Σ is diagonal: $\Sigma = \text{diag}(\sigma_i^2)$.
- Assuming Σ is *stationary*: $\Sigma_{ij} = \sigma(i-j)$ for some function $\sigma(\cdot)$.
- Assuming μ or Σ is sparse.

The contours of the multivariate Gaussian pdf $f_X(x)$ are given by the points $x \in \mathbb{R}^p$ such that

$$(x-\mu)^T \Sigma^{-1} (x-\mu) = c,$$

for some $c \in \mathbb{R}$. This can be interpreted via the principal components transformation $y = U^T(x-\mu)$ where $U\Lambda U^T = \Sigma$ is the spectral decomposition of (let's assume positive definite) Σ. Then, we have the above is equivalent to $y^T \Lambda^{-1} y = c \iff \sum_{i=1}^p \frac{y_i^2}{\lambda_i} = c$.

Thus, the contours of a multivariate Gaussian pdf are ellipses in $\mathbb{R}^p$.

The moment generating function of a multivariate Gaussian $X \sim \mathcal{N}_p(\mu, \Sigma)$ is

$$M_X(t) := \mathbb{E}[e^{t^T X}] = e^{t^T \mu} \cdot \mathbb{E}[e^{t^T \Sigma^{1/2} Z}] = e^{t^T \mu} \cdot \mathbb{E}[e^{(\Sigma^{1/2} t)^T Z}].$$

Now, $(\Sigma^{1/2} t)^T Z = a_1 Z_1 + \cdots + a_p Z_p$ is a linear combination of independent $\mathcal{N}(0,1)$ random variables. Then, $\mathbb{E}[e^{(\Sigma^{1/2} t)^T Z}]$ is the mgf of the random variable $(\Sigma^{1/2} t)^T Z$ evaluated at $s = 1$ or:

$$\mathbb{E}[e^{(\Sigma^{1/2} t)^T Z}] = M_{(\Sigma^{1/2} t)^T Z}(1) = \prod_{i=1}^{p} M_{a_i Z_i}(1) = \prod_{i=1}^{p} M_{Z_i}(a_i) = \prod_{i=1}^{p} e^{a_i^2/2} = e^{\sum_{i=1}^{p} a_i^2/2} = e^{\|\Sigma^{1/2} t\|_2^2/2} = e^{\frac{1}{2} t^T \Sigma t}.$$

Thus,

$$M_X(t) = e^{t^T \mu + \frac{1}{2} t^T \Sigma t}.$$

The mgf gives us another characterization of a multivariate Gaussian. In the above calculation, we essentially showed that the mgf of multivariate Gaussian X is the univariate mgf of a linear combination of Gaussian random variables.

Theorem 1.1 (Cramér-Wold device)

$X \sim \mathcal{N}_p(\mu, \Sigma)$ iff $a^T X \sim \mathcal{N}(a^T \mu, a^T \Sigma a)$ for all $a \in \mathbb{R}^p$.

Proof. This follows from the mgf formula above. ∎

Next, sums (and linear combinations) of multivariate Gaussians behave nicely just as in the one-dimensional case: if $X_1 \sim \mathcal{N}_p(\mu_1, \Sigma_1)$ and $X_2 \sim \mathcal{N}_p(\mu_2, \Sigma_2)$, then

$$X_1 + X_2 \sim \mathcal{N}_p(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2).$$

It's sometimes useful to think about a partition of a random vector X into two components $X = (X_1, X_2)$ where X_1, X_2 are random vectors of smaller dimension. In the Gaussian case, where $X_1 \sim \mathcal{N}_p(\mu_1, \Sigma_{11})$ and $X_2 \sim \mathcal{N}_p(\mu_2, \Sigma_{22})$, and X_1, X_2 are jointly Gaussian, we have

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \mathcal{N}_{p+q}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right). \tag{1}$$

The covariance matrix above is a block matrix where the off-diagonal block $\Sigma_{12} = \Sigma_{21}^T = \mathrm{Cov}(X_1, X_2)$.

One special property of the Gaussian family is that the covariance/correlation characterizes independence. We showed in the last review session that this is of course not true in general.

Theorem 1.2

Suppose $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ is jointly Gaussian. Then, X_1 and X_2 are independent iff $\mathrm{Cov}(X_1, X_2) = 0$.

Proof. This follows from the formula for the mgf of X. If $\mathrm{Cov}(X)$ is a block diagonal matrix, then the mgf factors, yielding independence. ∎

A linear transformation AX of a Gaussian random vector $X \sim \mathcal{N}_p(\mu, \Sigma)$ for $A \in \mathbb{R}^{q \times p}$ is again a Gaussian random vector by the Cramér-Wold device. The distribution is determined by the mean and covariance:

$$AX \sim \mathcal{N}_q(A\mu, A\Sigma A^T).$$

One particular kind of linear transformation is a projection. For example, AX can be a projection of X onto a subset of its p coordinates. By the above, we have that any marginal distribution of X must be Gaussian.

Theorem 1.3

If $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ has a joint normal distribution, then $X_1 \sim \mathcal{N}_p(\mu_1, \Sigma_{11})$ where μ_1, Σ_{11} can be read off from the mean and covariance of X.

> **Remark 1.4.** The converse of the above statement is not necessarily true. Two random variables X_1, X_2 may be marginally Gaussian, but not jointly Gaussian (see Problem 7).

Lastly, the conditional distribution of one Gaussian random vector conditioned on another is again Gaussian. If X_1, X_2 are jointly Gaussian as in (1), then

$$X_2 | X_1 = x \sim \mathcal{N}_q \left(\mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (x - \mu_1), \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right). \tag{2}$$

We'll give the general idea for the proof in the simpler case where X_1, X_2 are one-dimensional with $X_1 \sim \mathcal{N}(0, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(0, \sigma_2^2)$. Suppose $\rho := \mathrm{Cov}(X_1, X_2)$. Then, we have conditional density

$$\begin{aligned} f_{X_2|X_1=x_1}(x_2) &= \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)} \\ &\propto \frac{\exp\left(-\frac{1}{2(1-\rho^2)} \left(\left(\frac{x_1}{\sigma_1^2} \right)^2 - 2\rho \left(\frac{x_1}{\sigma_1} \right) \left(\frac{x_2}{\sigma_2} \right) + \left(\frac{x_2}{\sigma_2} \right)^2 \right) \right)}{\exp\left(-\frac{1}{2\sigma_2^2} x_2^2 \right)} \\ &\propto \exp\left(-\frac{1}{2} \left(a x_1^2 + b x_1 x_2 + c x_2^2 \right) \right), \end{aligned}$$

where a, b, c are some constants in terms of σ_1, σ_2, ρ. Next, we "complete the square" w.r.t. the variable x_2 in the quadratic above. Essentially, to get this into the form of a Gaussian pdf, we want to factor the quadratic as something resembling $\frac{(x_2-d)^2}{e}$ for some constants d, e. Fortunately, we can multiply the above formula by any constant not depending on x_2 since this does not change the kernel of the pdf (and only changes the normalizing constant). Thus, we have

$$f_{X_2|X_1=x_1}(x_2) \propto \exp\left(-\frac{1}{2} \left(a x_1^2 + b x_1 x_2 + c x_2^2 + \left(\frac{b^2}{4c} - a \right) x_1^2 \right) \right) = \exp\left(-\frac{1}{2/c} (x_2 + x_1 b/(2c))^2 \right).$$

Thus, we have shown $X_2 | X_1 = x_1$ is Gaussian. Its mean and variance can be given in terms of σ_1, σ_2, ρ by following the calculations above carefully. The proof of (2) for multivariate Gaussians is similar and also involves a "completing the square" trick.

2 Properties of a Random Sample

Often, the data collected in an experiment consist of several observations on a variable of interest (e.g., the height of persons drawn at random from a population).

We say $X_1, \dots, X_n$ are a *random sample* of size n from a population density $f(x)$ if $X_1, \dots, X_n$ are mutually independent random variables and the marginal pdf/pmf of each X_i is the same function $f(x)$. Another of way of saying this is that $X_1, \dots, X_n$ are *independent and identically distributed* random variables with pdf/pmf $f(x)$. This is often just abbreviated as i.i.d., or $\{X_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} f$.

From a sample $\{X_1, \dots, X_n\}$, we might want to obtain some summary of the values within a sample. Formally, this is a function $T : \mathbb{R}^n \to \mathbb{R}$ taking as inputs $X_1, \dots, X_n$. We call the random variable T a *statistic* and refer to its distribution as a *sampling distribution*. The two most standard examples of a statistic are:

1. The *sample mean*: $\overline{X} := \frac{1}{n} \sum_{i=1}^n X_i$.
2. The *sample variance*: $S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2$.

These are estimates or guesses for the population mean and variance, respectively, based on the sample $\{X_1, \dots, X_n\}$. They exhibit several universal properties and are particularly well-behaved in the Gaussian setting (i.e., when f is normal). Let's first derive some generic properties.

Notation 2.1. Let $\overline{x}, s^2, s$ denote observed values of the random variables $\overline{X}, S^2, S := \sqrt{S^2}$.

Theorem 2.2

Let $x_1, \ldots, x_n \in \mathbb{R}$. Then,

(i) $\min_a \sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \overline{x})^2$

(ii) $(n-1)s^2 = \sum_{i=1}^n (x_i - \overline{x})^2 = \sum_{i=1}^n x_i^2 - n\overline{x}^2$

The proof is very similar to the analogous statements in the population setting (see Example 8 in Review Doc 2). Hint: (i) is proven by adding and subtracting $\overline{x}$ inside each square and then expanding the square.

Theorem 2.3

Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Then

(i) $\mathbb{E}[\overline{X}] = \mu$.

(ii) $\operatorname{Var}(\overline{X}) = \sigma^2/n$.

(iii) $\mathbb{E}[S^2] = \sigma^2$.

Proof. (i) and (ii) follow from the linearity properties of the expectation and variance, respectively. (iii) follows from linearity of expectation and Theorem 2.2. ∎

The above theorem gives us a relationship between a statistic and the population parameter it represents. In particular, (i) and (iii) show that $\overline{X}$ and S^2 are *unbiased estimators* of their parameters μ and σ^2, respectively. In general, we say $T(X_1, \ldots, X_n)$ is unbiased if $\mathbb{E}[T(X_1, \ldots, X_n)]$ is equal to the parameter it represents. Typically, the parameter is first identified and then the estimator T is proposed for determining its value based on the sample.

2.1 Sampling from a Normal Distribution

When $X_1, \ldots, X_n$ are sampled from a Gaussian distribution, the sample quantities $\overline{X}$ and S^2 exhibit additional useful properties. We start with two lemmas which will help us better understand the relationship between $\overline{X}$ and S^2.

Lemma 2.4 (chi squared and Gaussians)

Recall χ_p^2 denotes a chi squared random variable with p degrees of freedom.

1. If $Z \sim \mathcal{N}(0,1)$, then $Z^2 \sim \chi_1^2$.
2. If $X_1, \ldots, X_n$ are independent and $X_i \sim \chi_p^2$, then $X_1 + \cdots + X_n \sim \chi^2_{p_1 + \cdots + p_n}$.

Proof. The first part can be deduced from the pdf transformation law. The second part follows from an mgf computation, where the mgf of each chi-squared X_i is $(1-2t)^{-p_i/2}$. ∎

Lemma 2.5

Let $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 1, \dots, n$, independent. For constants a_{ij}, b_{rj} ($j = 1, \dots, n; i = 1, \dots, k; r = 1, \dots, m$), where $k + m \le n$, define

$$U_i = \sum_{j=1}^n a_{ij} X_j, i = 1, \dots, k$$

$$V_r = \sum_{j=1}^n b_{rj} X_j, r = 1, \dots, m$$

Then,

1. The random variables U_i, V_r are independent iff $\mathrm{Cov}(U_i, V_r) = 0$. Furthermore, $\mathrm{Cov}(U_i, V_r) = \sum_{j=1}^n a_{ij} b_{rj} \sigma_j^2$,
2. The random vectors $(U_1, \dots, U_k)$ and $(V_1, \dots, V_m)$ are independent iff U_i is independent of V_r for all pairs i, r ($i = 1, \dots, k; r = 1, \dots, m$).

Proof. Both of these follow from the Cramér-Wold device. ∎

Theorem 2.6

Let $X_1, \dots, X_n$ be a random sample from a $\mathcal{N}(\mu, \sigma^2)$ distribution. Then,

(i) $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.

(ii) $\overline{X}$ and S^2 are independent random variables.

(iii) $(n-1)S^2/\sigma^2$ has a chi squared distribution with $n-1$ degrees of freedom.

Proof. (i) is clear from properties of a Gaussian. We will show (ii) by showing that S^2 can be represented as some function of the random vector $(X_2 - \overline{X}, \dots, X_n - \overline{X})$. Then, it suffices to show $(X_2 - \overline{X}, \dots, X_n - \overline{X})$ and $\overline{X}$ are independent. First, we have

$$S^2 = \frac{1}{n-1}\left((X_1 - \overline{X})^2 + \sum_{i=2}^n (X_i - \overline{X})^2\right) = \frac{1}{n-1}\left(\left(\sum_{i=2}^n (X_i - \overline{X})\right)^2 + \sum_{i=2}^n (X_i - \overline{X})^2\right).$$

The second equality is established by using the identity $\sum_{i=1}^n (X_i - \overline{X}) = 0$. Now, let $Y_1 := \overline{X}$ and for $i = 2, \dots, n$, let $Y_i := X_i - \overline{X}$. Then, $S^2 = g(Y_2, \dots, Y_n)$ is some function of $Y_1, \dots, Y_n$. We then claim Y_1 and $(Y_2, \dots, Y_n)$ are independent. In fact, by Lemma 2.5, it suffices to show Y_1 is uncorrelated with each Y_i for $i = 2, \dots, n$. Indeed, for $j > 1$:

$$\mathrm{Cov}(\overline{X}, X_j - \overline{X}) = \sum_{i=1}^n \left(\frac{1}{n}\right) \cdot \left(\mathbf{1}\{i = j\} - \frac{1}{n}\right) = 0.$$

Thus, S^2 and $\overline{X}$ are independent.

To show (iii), we first assume without loss of generality that each $X_i \sim \mathcal{N}(0, 1)$. This is allowed since the value of $S^2/\sigma^2 = \frac{1}{\sigma^2(n-1)} \sum_{i=1}^n (X_i - \overline{X})^2$ does not change under the transformation $(X_1, \dots, X_n) \mapsto \left(\frac{X_1 - \mu}{\sigma}, \dots, \frac{X_n - \mu}{\sigma}\right)$. More precisely, we have

$$S^2/\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \overline{Z})^2,$$

where $(Z_1, \dots, Z_n) := \left(\frac{X_1 - \mu}{\sigma}, \dots, \frac{X_n - \mu}{\sigma}\right)$ and $\overline{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$ are the standardized sample and standardized sample mean, respectively.

Thus, it remains to show $\sum_{i=1}^n (X_i - \overline{X})^2 \sim \chi^2_{n-1}$ for $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. We have

$$(n-1)S^2 = \sum_{i=1}^n X_i^2 - n\overline{X}^2 \implies (n-1)S^2 + n\overline{X}^2 = \sum_{i=1}^n X_i^2.$$

Taking the mgf of both sides of the above and using (ii), we have

$$M_{\sum_{i=1}^n X_i^2}(t) = M_{(n-1)S^2}(t) \cdot M_{n\overline{X}^2}(t).$$

Now, $\sqrt{n}\cdot\overline{X} \sim \mathcal{N}(0,1)$. Thus, from Lemma 2.4, $n\overline{X}^2 \sim \chi^2_1$. Additionally, Lemma 2.4 also gives us $\sum_{i=1}^n X_i^2 \sim \chi^2_n$. Thus, using the fact that mgf of a $V \sim \chi^2_k$ distribution is $M_V(t) = (1-2t)^{-k/2}$, we have that $M_{(n-1)S^2}(t) = (1-2t)^{-(n-1)/2}$ meaning $(n-1)S^2 \sim \chi^2_{n-1}$. ■

If $X_1, \ldots, X_n$ are a random sample from $\mathcal{N}(\mu, \sigma^2)$, then the *standardized mean*

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}},$$

is distributed as a $\mathcal{N}(0,1)$ random variable. If we knew the value of σ and measured $\overline{X}$, we could use the standardized mean as a basis for inference about μ, since μ would then be the only unknown quantity. However, it is often the case that μ is unknown. This leads us to the *Student's t distribution*:

$$\frac{\overline{X} - \mu}{S/\sqrt{n}}.$$

The distribution of this random variable appears at first glance complicated. However, since we know $\overline{X}$ and S^2 are independent, the Student's t distribution is really a ratio of two independent random variables.

Definition 2.7 (Student's t distribution)**.** Let $X_1, \ldots, X_n$ be a random sample from a $\mathcal{N}(\mu, \sigma^2)$ distribution. The quantity $(\overline{X} - \mu)/(S/\sqrt{n})$ has a *Student's t distribution* with $n-1$ degrees of freedom. Equivalently, a random variable T has a Student's t distribution with p degrees of freedom and we write $T \sim t_p$, if it has pdf

$$f_T(t) = \frac{\Gamma\left(\frac{p-1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \cdot \frac{1}{(p\pi)^{1/2}} \cdot \frac{1}{(1+t^2/p)^{(p+1)/2}}, -\infty < t < \infty$$

For $p = 1$, this becomes the pdf of the Cauchy distribution.

> **Remark 2.8.** The Student's t has no mgf because it does not have moments of all orders. In fact, if there are p degrees of freedom, then there are only $p-1$ moments. Hence, a t_1 distribution has no mean, a t_2 has no variance, etc. If $T_p \sim t_p$, then
>
> $$\mathbb{E}[T_p] = 0 \text{ if } p > 1 \text{ and } \operatorname{Var} T_p = \frac{p}{p-2} \text{ if } p > 2$$

Example 2.9 (variance ratio distribution)

Let $X_1, \ldots, X_n$ be a random sample from a $\mathcal{N}(\mu_X, \sigma_X^2)$ population, and let $Y_1, \ldots, Y_m$ be a random sample from an independent $\mathcal{N}(\mu_Y, \sigma_Y^2)$ population. Consider the ratio σ_X^2/σ_Y^2. Information about this ratio is contained in S_X^2/S_Y^2, the ratio of sample variances. The F distribution allows us to compare these quantities by giving us a distribution of

$$\frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$$

Definition 2.10 (F distribution)**.** Under the same setup as the previous example, the random variable

$$F := (S_X^2/\sigma_X^2)/(S_Y^2/\sigma_Y^2)$$

has *Snedecor's F distribution* with $n-1$ and $m-1$ degrees of freedom. Equivalently, the random variable F has the F distribution with p and q degrees of freedom if it has pdf

$$f_F(x) = \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \left(\frac{p}{q}\right)^{p/2} \cdot \frac{x^{(p/2)-1}}{(1+(p/q)x)^{(p+q)/2}}, 0 < x < \infty$$

Theorem 2.11

We have

1. If $X \sim F_{p,q}$, then $1/X \sim F_{q,p}$; that is, the reciprocal of an F random variable is again an F random variable.
2. If $X \sim t_q$, then $X^2 \sim F_{1,q}$.
3. If $X \sim F_{p,q}$, then $(p/q)X/(1+(p/q)X) \sim$ beta$(p/2, q/2)$.

2.2 Order Statistics

Definition 2.12 (order statistics)**.** The *order statistics* of a random sample $X_1, \dots, X_n$ are the sample values placed in ascending order, i.e. $X_{(1)} = \min_i X_i$, $X_{(2)}$ is the second smallest X_i, and so on.

Theorem 2.13

Let $X_{(1)}, \dots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \dots, X_n$, from a contnuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the pdf of $X_{(j)}$ is

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) F_X(x)^{j-1}(1-F_X(x))^{n-j}$$

Example 2.14 (uniform order statistics pdf)

Let $X_1, \dots, X_n$ be iid uniform$(0,1)$. Using the previous result, we have that the pdf of the j-th order statistic is

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} x^{j-1}(1-x)^{n-j} = \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} x^{j-1}(1-x)^{(n-j+1)-1} \text{ for } x \in (0,1)$$

Thus, the j-th order statistic from a uniform$(0,1)$ sample has a beta$(j, n-j+1)$ distribution. Thus,

$$\mathbb{E}[X_{(j)}] = \frac{j}{n+1} \text{ and } \operatorname{Var}(X_{(j)}) = \frac{j(n-j+1)}{(n+1)^2(n+2)}$$

Theorem 2.15

Let $X_{(1)}, \dots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \dots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the joint pdf of $X_{(i)}$ and $X_{(j)}$, $1 \le i < j \le n$, is

$$f_{X_{(i)},X_{(j)}}(u,v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(u) f_X(v) F_X(u)^{i-1}(F_X(v)-F_X(u))^{j-1-i}(1-F_X(v))^{n-j}$$

for $-\infty < u < v < \infty$. The joint pdf of all the order statistics is given by

$$f_{X_{(1)},\dots,X_{(n)}}(x_1, \dots, x_n) = \begin{cases} n! f_X(x_1) \cdots f_X(x_n) & -\infty < x_1 < \cdots < x_n < \infty \\ 0 & \text{otherwise} \end{cases}$$

3 Problems

3.1 Previous Core Competency Problems

Problem 1. [2018 Summer Practice, # 10] Suppose that $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(0,1)$, and A is an $n \times n$ matrix which is symmetric (i.e., $A^T = A$) and idempotent (i.e., $A^2 = A$). Find the distribution of $\sum_{i,j=1}^{n} X_i X_j A(i,j)$. Assume if necessary that $\sum_{i=1}^{n} A(i,i) = s$.

Problem 2 (2018 Summer Practice, # 17)**.** Let X and Y be i.i.d. $\mathcal{N}(0,1)$ random variables. Consider

$$Z := \text{sign}(Y) \cdot X$$

where sign$(y) := 1$ if $y > 0$ and sign$(y) := -1$ if $y \leq 0$.

(a) Find the distribution of Z.

(b) Compute the covariance of X and Z.

(c) Determine $P[X + Z = 0]$.

(d) Are X and Z independent? (Give a precise mathematical argument).

Problem 3 (2018 September, # 8)**.** Suppose $(\mathbf{X}, \mathbf{Y})$ have a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix

$$\Sigma = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix},$$

where A is $m \times m$, B is $m \times n$, and C is $n \times n$, and A and C are non-singular. Define a vector $\mathbf{Z} := \mathbf{Y} - B^T A^{-1} \mathbf{X}$.

(i) Find the $m \times n$ covariance matrix of $\mathbf{X}$ and $\mathbf{Z}$.

(ii) Express $\mathbf{Y}$ as $\mathbf{Z} + B^T A^{-1} \mathbf{X}$, and, hence deduce the conditional distribution of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$.

Problem 4. [2018 September, # 9] Let $X \in \mathbb{R}^d$ be a centered normal random vector and $A \in \mathbb{R}^{d \times d}$ a fixed symmetric matrix. Denote by Y an independent copy of X. Show that

$$X^T A X - Y^T A Y \overset{d}{=} 2 X^T A Y.$$

Hint: $(X \pm Y)/\sqrt{2}$ are i.i.d. random vectors following the same distribution as X.

Problem 5 (2020 September, # 7)**.** Suppose X_1, X_2 are i.i.d. $N(0,1)$.

(a) Find the joint distribution of $X_1 + X_2$ and $X_1 - X_2$.

(b) Show that $2X_1X_2$ has the same distribution as $X_1^2 - X_2^2$.

3.2 Additional Practice

Problem 6 (Casella & Berger, Exercise 4.20)**.** Suppose X_1, X_2 are independent $\mathcal{N}(0, \sigma^2)$ random variables.

1. Find the joint distribution of Y_1 and Y_2, where

$$Y_1 = X_1^2 + X_2^2 \text{ and } Y_2 = \frac{X_1}{\sqrt{Y_1}}.$$

2. Show that Y_1 and Y_2 are independent, and interpret this result geometrically.

Problem 7 (marginal normality does not imply bivariate normality)**.** [Casella & Berger, Exercise 4.47] Let X and Y be independent $\mathcal{N}(0,1)$ random variables, and define a new random variable Z by

$$Z = \begin{cases} X & XY > 0 \\ -X & XY < 0 \end{cases}.$$

1. Show that Z has a normal distribution.

2. Show that the joint distribution of Z and Y is not bivariate normal. Hint: show that Z and Y always have the same sign.

Problem 8 (Casella & Berger, Exercise 5.22)**.** Let X and Y be iid $\mathcal{N}(0,1)$ random variables, and define $Z = \min(X, Y)$. Prove that $Z^2 \sim \chi_1^2$.