

# STAT151A Homework 3: Due February 23rd

Your name here

Normal intervals

For these problems, assume I give you a computer program that can compute the function  $\Phi(z) = \mathbb{P}(\tilde{z} \leq z)$  where  $\tilde{z}$  is a standard scalar-valued random variable.

Let  $\tilde{x}$  denote a scalar-valued  $N(\mu, \sigma^2)$  random variable. Using only  $\Phi(z)$  and elementary arithmetic, construct functions that evaluate the following:

**(a)**

$$a \mapsto \mathbb{P}(\tilde{x} \leq a)$$

**A solution:**

$$\mathbb{P}(\tilde{x} \leq a) = \mathbb{P}\left(\frac{\tilde{x} - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) = \mathbb{P}\left(\tilde{z} \leq \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right)$$

**(b)**

$$b \mapsto \mathbb{P}(\tilde{x} \geq b)$$

**A solution:**

$$\mathbb{P}(\tilde{x} \geq b) = \mathbb{P}\left(\frac{\tilde{x} - \mu}{\sigma} \geq \frac{b - \mu}{\sigma}\right) = \mathbb{P}\left(\tilde{z} \geq \frac{b - \mu}{\sigma}\right) = \mathbb{P}\left(\tilde{z} \leq -\frac{b - \mu}{\sigma}\right) = \Phi\left(-\frac{b - \mu}{\sigma}\right)$$

**(c)**

$$a, b \mapsto \mathbb{P}(b \leq \tilde{x} \leq a)$$

**A solution:**

$$\mathbb{P}(\tilde{x} \leq a) = \mathbb{P}(b \leq \tilde{x} \leq a) + \mathbb{P}(\tilde{x} \leq b) \quad \Rightarrow \quad \mathbb{P}(b \leq \tilde{x} \leq a) = \mathbb{P}(\tilde{x} \leq a) - \mathbb{P}(\tilde{x} \leq b)$$

and use (a).

**(d)**

$$a \mapsto \mathbb{P}(|\tilde{x}| \leq a)$$

**A solution:**

$$\mathbb{P}(|\tilde{x}| \leq a) = \mathbb{P}(-a \leq \tilde{x} \leq a) \text{ and use (c).}$$

**(e)**

$$a \mapsto \mathbb{P}(|\tilde{x}| \geq a)$$

**A solution:**

$$\mathbb{P}(|\tilde{x}| \geq a) = 1 - \mathbb{P}(|\tilde{x}| \leq a) \text{ and use (d).}$$

**(f)**

$$a \mapsto \mathbb{P}(|\tilde{x}| > a)$$

**A solution:**

$$\tilde{x} \text{ is continuous, so } \mathbb{P}(\tilde{x} = a) = 0 \text{ and } \mathbb{P}(|\tilde{x}| > a) = \mathbb{P}(|\tilde{x}| \geq a). \text{ Use (e).}$$

**(g)**

$$a \mapsto \mathbb{P}(|\tilde{x}| = a)$$

**A solution:**  $\mathbb{P}(|\tilde{x}| = a) = 0$ .

Multivariate CLT

Let  $\tilde{\mathbf{x}}_n$  denote an IID sequence of random variables in  $\mathbb{R}^P$  (not necessarily normal), each with zero mean and finite covariance matrix  $\mathbf{\Sigma}$ . Let  $\mathbf{a} \in \mathbb{R}^P$  denote a fixed vector.

**(a)**

Using the univariate CLT, find the limiting distribution of

$$\frac{1}{\sqrt{N}} \sum_{n=1}^N \mathbf{a}^\top \tilde{\mathbf{x}}_n.$$

**Solution:**

Since  $\mathbf{a}$  is constant,  $\mathbb{E}[\mathbf{a}^\top \tilde{\mathbf{x}}_n] = 0$ , so we can apply the CLT. We also have  $\text{Cov}(\mathbf{a}^\top \tilde{\mathbf{x}}_n) = \mathbf{a}^\top \Sigma \mathbf{a}$ , so the limiting distribution is  $\mathcal{N}(0, \mathbf{a}^\top \Sigma \mathbf{a})$ .

**(b)**

Using the multivariate CLT and the continuous mapping theorem, find the limiting distribution of

$$\mathbf{a}^\top \left( \frac{1}{\sqrt{N}} \sum_{n=1}^N \tilde{\mathbf{x}}_n \right).$$

**Solution:** By the multivariate CLT,  $\frac{1}{\sqrt{N}} \sum_{n=1}^N \tilde{\mathbf{x}}_n \rightarrow \tilde{\mathbf{z}}$  where  $\tilde{\mathbf{z}} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ . By the continuous mapping theorem, the distribution of the given expression converges to the distribution of  $\mathbf{a}^\top \tilde{\mathbf{z}}$ , which is  $\mathcal{N}(0, \mathbf{a}^\top \Sigma \mathbf{a})$  as in (a). (As must be the case.)

**(c)**

Now, suppose that  $P = 2$  and

$$\Sigma = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Note that we can write

$$\tilde{\mathbf{x}}_n = \begin{pmatrix} \tilde{\mathbf{x}}_{n1} \\ \tilde{\mathbf{x}}_{n2} \end{pmatrix},$$

where  $\tilde{\mathbf{x}}_{n1}$  and  $\tilde{\mathbf{x}}_{n2}$  are scalars. Find the limiting distributions of each of the following expressions:

$$\begin{aligned}\frac{1}{\sqrt{N}} \sum_{n=1}^N \tilde{\mathbf{x}}_{n1} &\rightarrow? \\ \frac{1}{\sqrt{N}} \sum_{n=1}^N \tilde{\mathbf{x}}_{n2} &\rightarrow? \\ \frac{1}{\sqrt{N}} \sum_{n=1}^N (\tilde{\mathbf{x}}_{n1} + \tilde{\mathbf{x}}_{n2}) &\rightarrow?\end{aligned}$$

(This result demonstrates why it's not enough to only look at the marginal distribution of the vector components when using a multivariate CLT.)

**Solution:** The first two go to  $\mathcal{N}(0, 1)$ , but the last goes to  $\mathcal{N}(0, 0)$ , since

$$\tilde{\mathbf{x}}_{n1} + \tilde{\mathbf{x}}_{n2} = (1, 1)\tilde{\mathbf{x}} \quad \Rightarrow \quad \text{Var}(\tilde{\mathbf{x}}_{n1} + \tilde{\mathbf{x}}_{n2}) = (1, 1) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0.$$

Valid covariance matrices

Suppose I were to tell you that the vector-valued random variable  $\mathbf{x}$  has a covariance matrix  $\text{Cov}(\mathbf{x}) = \mathbf{\Sigma}$  where  $\mathbf{\Sigma}$  is not positive semi-definite (i.e.,  $\mathbf{\Sigma}$  has at least one negative eigenvalue). Show that, if this were true, you could construct a scalar-valued random variable with *negative* variance, which is impossible.

(It follows from this argument every covariance matrix must be positive semi-definite.)

**Solution:**

Let  $\mathbf{u}$  denote an eigenvector of  $\mathbf{\Sigma}$  with a negative eigenvalue. Then  $\text{Var}(\mathbf{u}^\top \mathbf{x}) = \mathbf{u}^\top \mathbf{\Sigma} \mathbf{u} < 0$ .