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 $\Phi(z) = \mathbb{P}\left(\tilde{z} \le z\right)$ 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)

$a \mapsto \mathbb{P}\left(\tilde{x} \le a\right)$

A solution:

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

(b)

$b \mapsto \mathbb{P}\left(\tilde{x} \ge b\right)$

A solution:

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

(c)

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

A solution:

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

and use (a).

(d)

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

A solution:

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

(e)

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

A solution:

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

(f)

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

A solution:

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

(g)

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

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

Multivariate CLT

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

(a)

Using the univariate CLT, find the limiting distribution of

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

Solution:

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

(b)

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

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

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

(c)

Now, suppose that $P = 2$ and

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

Note that we can write

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

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

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \tilde{\boldsymbol{x}}_{n1} \rightarrow ?$$

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \tilde{\boldsymbol{x}}_{n2} \rightarrow ?$$

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{N} (\tilde{\boldsymbol{x}}_{n1} + \tilde{\boldsymbol{x}}_{n2}) \rightarrow ?$$

(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

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

Valid covariance matrices

Suppose I were to tell you that the vector-valued random variable $\boldsymbol{x}$ has a covariance matrix $\mathrm{Cov}\left(\boldsymbol{x}\right) = \boldsymbol{\Sigma}$ where $\boldsymbol{\Sigma}$ is not positive semi-definite (i.e., $\boldsymbol{\Sigma}$ 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 postive semi-definite.)

Solution:

Let $\boldsymbol{u}$ denote an eigenvector of $\boldsymbol{\Sigma}$ with a negative eigenvalue. Then $\mathrm{Var}\left(\boldsymbol{u}^\intercal \boldsymbol{x}\right) = \boldsymbol{u}^\intercal \boldsymbol{\Sigma} \boldsymbol{u} < 0$.