1. Problem 1

(a) "If side"

We use characteristic function.

First, let $Y = a^T X$, then the characteristic function of Y is

$$\varphi_Y(t) = \mathbb{E}(e^{itY}) = \mathbb{E}(e^{it \cdot a^T X}), \ \forall t, \ \forall a \in \mathbb{R}^m$$

It's known that $Y = a^T X$ is normal distribution, so

$$\varphi_Y(t) = e^{it \cdot \mathbb{E}(Y) - \frac{1}{2} var(Y) t^2} = \mathbb{E}(e^{it \cdot a^T X}), \forall t, \forall a \in \mathbb{R}^m$$

We can take $t = 1$, then

$$\mathbb{E}(e^{ia^T X}) = e^{i \mathbb{E}(Y) - \frac{1}{2} var(Y)}, \forall a \in \mathbb{R}^m$$

Note that $Y = a^T X$, then $\mathbb{E}(Y) = a^T \mathbb{E}(X) = a^T \mu$.

$$var(Y) = a^T var(X) a = a^T \Sigma a$$

And the characteristic function of X is

$$\varphi_X(t) = \mathbb{E}(e^{it^T X}), \forall t \in \mathbb{R}^m$$

From above, we have

$$\varphi_X(t) = e^{it^T \mu - \frac{1}{2} t^T \Sigma t}, \ \mu = \mathbb{E}(X), \ \Sigma = var(X)$$

Note that this form is identical to the characteristic function of multivariate normal distribution $N(\mu, \Sigma)$. By the property of characteristic function, X follows $N(\mu, \Sigma)$

"Only if" side.

From above, if $X \sim N(\mu, \Sigma)$, then

$$\varphi_X(t) = e^{it^T \mu - \frac{1}{2} t^T \Sigma t}, \ t \in \mathbb{R}^m$$

We can always write t as $t = ka$, where $a \in \mathbb{R}$, $k \in \mathbb{R}^m$

then, $\varphi_X(t) = \mathbb{E}(e^{it^TX}) = \mathbb{E}(e^{ika^TX})$

$$= \varphi_{a^TX}(k) = e^{ik(a^T\mu) - \frac{1}{2}(a^T\Sigma a)\cdot k^2}$$

$$= e^{ik\,\mathbb{E}(a^TX) - \frac{1}{2}\mathrm{var}(a^TX)\cdot k^2}, \quad \forall a \in \mathbb{R},\ k \in \mathbb{R}^m$$

Note that the characteristic function of a^TX is identical to that of $N(a^T\mu, a^T\Sigma a)$.

Therefore, $\forall a \in \mathbb{R}^m$, the linear combination a^TX is normally distributed.

(b)

"If" side:

We have Σ is a real symmetric matrix, then by spectrum decomposition, $\Sigma = Q\Lambda Q^T$, where Q is an orthogonal matrix and Λ is a diagonal matrix whose entries are eigenvalues of Σ.

Since we have $\det(\Sigma) > 0$, then all elements of Λ's diagonal are positive. So we can write $\Lambda = \Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}$, such that $\Lambda^{\frac{1}{2}}$ have diagonal entries of squared root of Λ's diagonal. Then

$$\Sigma = Q\Lambda Q^T = (Q\Lambda^{\frac{1}{2}})(Q\Lambda^{\frac{1}{2}})^T = \Sigma^{\frac{1}{2}}(\Sigma^{\frac{1}{2}})^T,$$

$$\Sigma^{\frac{1}{2}} \in \mathbb{R}^{m\times m}$$

Now, by definition of multivariate normal distribution, we have $X = \mu + \Sigma^{\frac{1}{2}}Z$, $Z \sim N(0, \mathbb{I}_m)$

We have the pdf of Z is.

$$f_Z(\vec{z}) = \frac{1}{(\sqrt{2\pi})^m} \cdot \exp\left[-\frac{1}{2}\sum_{i=1}^{m} z_i^2\right]$$

$$= \frac{1}{(\sqrt{2\pi})^m} \exp\left[-\frac{1}{2}\vec{z}^T\vec{z}\right]$$

From $X = \mu + \Sigma^{\frac{1}{2}} Z$, $\left|\frac{\partial \vec{x}}{\partial \vec{z}}\right| = \left|\Sigma^{\frac{1}{2}}\right|$.

Moreover $\Sigma^{\frac{1}{2}} = Q\Lambda^{\frac{1}{2}}$, we define $\Lambda^{-\frac{1}{2}}$ as diagonal taking the reciprocal of $\Lambda^{\frac{1}{2}}$, then we have $\Lambda^{\frac{1}{2}}\Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}}\Lambda^{\frac{1}{2}} = \mathbb{I}_m$. Therefore, we find out the inverse of $\Sigma^{\frac{1}{2}}$, denote by $\Sigma^{-\frac{1}{2}}$, and $\Sigma^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}}Q^T$.

So $Z = \Sigma^{-\frac{1}{2}}(X-\mu)$. Then pdf of X is

$$f_X(\vec{x}) = \frac{1}{(\sqrt{2\pi})^m} \exp\left\{-\frac{1}{2}(x-\mu)^T\left(\Sigma^{-\frac{1}{2}}\right)^T\Sigma^{-\frac{1}{2}}(x-\mu)\right\} \cdot \left|\Sigma^{\frac{1}{2}}\right|^{-1}$$

Note that $\left(\Sigma^{-\frac{1}{2}}\right)^T\Sigma^{-\frac{1}{2}} = Q\Lambda^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}Q^T = Q\Lambda^{-1}Q^T$ which is inverse of Σ, denote by Σ^{-1}, and

$$|\Sigma| = \left|\Sigma^{\frac{1}{2}}\right|\left|\left(\Sigma^{\frac{1}{2}}\right)^T\right| = \left|\Sigma^{\frac{1}{2}}\right|^2 \rightarrow \left|\Sigma^{\frac{1}{2}}\right|^{-1} = |\Sigma|^{-\frac{1}{2}}$$

Therefore

$$f_X(\vec{x}) = \frac{1}{(2\pi)^{m/2}\sqrt{\det(\Sigma)}} \cdot \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\} \quad x \in \mathbb{R}^m$$

This pdf is obviously in $\mathbb{R}^m$ since we write $X = \mu + \Sigma^{\frac{1}{2}} Z$, and $\Sigma^{\frac{1}{2}} = Q\Lambda^{\frac{1}{2}}$ is well-defined.

And in this proof we definitely require $|\Sigma|>0$, since we require terms like Σ^{-1}, $\Sigma^{-\frac{1}{2}}$ and $|\Sigma|^{-1}$

"Only if" side:

Σ is the covariance matrix of X, so Σ is always positive semi-definite. So we must have $\det(\Sigma)>0$, or $\det(\Sigma)=0$.

If $\det(\Sigma)=0$, there must exist some vector $a \in \mathbb{R}^m$ such that $a^T\Sigma a=0$. Then we have

$$\mathbb{E}[a^T(x-\mu)(x-\mu)^T a] = a^T\mathbb{E}((x-\mu)(x-\mu)^T)a$$

$$= a^T \operatorname{cov}(X)\, a = 0$$

However, $\mathbb{E}[a^T(x-\mu)(x-\mu)^T a] = \mathbb{E}[\|(x-\mu)^T a\|_2^2] = 0$

Then term $\|(x-\mu)^T a\|_2^2$ is always non-negative.

Therefore to make the expectation be zero, we must have $(x-\mu)^T a = 0$. With this constraint, X can only take values in a hyperplane of dimension $m-1$, since if we choose the value of $x_1 \cdots x_{m-1}$, we can solve x_m from the equation above. Therefore, X can't has a pdf in $\mathbb{R}^m$ in this case.

Hence, we must have $\det(\Sigma)>0$.

The PDF of X is already calculated in "If" side

(c) Since $W \sim N(\mu_W, \Sigma_W)$, the MGF of W is

$$M_W(t) = \exp\left\{ t^T \mu_W + \frac{1}{2} t^T \Sigma_W t \right\} \quad t \in \mathbb{R}^{m+n}$$

We divide the vector t to two parts, $t = \begin{pmatrix} t_x \\ t_y \end{pmatrix}$ s.t.

$t_x \in \mathbb{R}^m$, $t_y \in \mathbb{R}^n$. then

$$M_W(t) = \exp\left\{ t_x^T \mu_x + t_y^T \mu_y + \frac{1}{2}(t_x^T \ t_y^T) \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix} \begin{bmatrix} t_x \\ t_y \end{bmatrix} \right\}$$

$$= \exp\left\{ t_x^T \mu_x + t_y^T \mu_y + \frac{1}{2}\left[t_x^T \Sigma_X t_x + t_y^T \Sigma_{YX} t_x \right.\right.$$

$$\left.\left. + t_x^T \Sigma_{XY} t_y + t_y^T \Sigma_Y t_y \right] \right\}$$

$$= \exp\left\{ t_x^T \mu_X + \frac{1}{2} t_x^T \Sigma_X t_x \right\} \cdot \exp\left\{ t_y^T \mu_Y + \frac{1}{2} t_y^T \Sigma_Y t_y \right\}$$

$$\cdot \exp\left\{ \frac{1}{2} t_y^T \Sigma_{YX} t_x + \frac{1}{2} t_x^T \Sigma_{XY} t_y \right\}$$

$$= \exp\left\{ t_x^T \mu_X + \frac{1}{2} t_x^T \Sigma_X t_x \right\} \cdot \exp\left\{ t_y^T \mu_Y + \frac{1}{2} t_y^T \Sigma_Y t_y \right\}$$

$$\cdot \exp\left\{ t_x^T \Sigma_{XY} t_y \right\} \quad \left[\text{since } \Sigma_{XY} = \Sigma_{YX}^T \right]$$

"If part":

If $\Sigma_{XY} = (\Sigma_{YX})^T = 0$, we have

$$M_W(t) = \exp\left\{ t_x^T \mu_X + \frac{1}{2} t_x^T \Sigma_X t_x \right\} \cdot \exp\left\{ t_y^T \mu_Y + \frac{1}{2} t_y^T \Sigma_Y t_y \right\}$$

Note that, the first exponential term is the MGF $N(\mu_X, \Sigma_X)$ and the second is $N(\mu_Y, \Sigma_Y)$

Now let the random vector $\tilde{W} = (\tilde{X}, \tilde{Y})$ s.t. $\tilde{X}$ and $\tilde{Y}$ are independent and $\tilde{X} \overset{d.}{=} X \sim N(\mu_X, \Sigma_X)$, $\tilde{Y} \overset{d.}{=} Y \sim N(\mu_Y, \Sigma_Y)$

We can always do this since we can just multiply $f_{\tilde{X}}(\tilde{x})$ and $f_{\tilde{Y}}(\tilde{y})$ to get the distribution of $f_{\tilde{W}}(\tilde{w})$.

Then, we find that $\tilde{W}$ and W have identical MGF.

Therefore $W \overset{d}{=} \tilde{W}$. Moreover, we have $\tilde{X} \overset{d}{=} X$ and $\tilde{Y} \overset{d}{=} Y$.

So we can divide $f_W(w)$ to multiplication of $f_X(x)$ and $f_Y(y)$. i.e. $f_W(w) = f_X(x) f_Y(y)$.

Therefore, X and Y are independent.

"Only if" part

From above, it's easy to get if X and Y are independent.

$$M_W(t) = \exp\{t_x^T \mu_X + \frac{1}{2} t_x^T \Sigma_X t_x\} \cdot \exp\{t_y^T \mu_Y + \frac{1}{2} t_y^T \Sigma_Y t_y\} \cdot \exp\{t_x^T \Sigma_{XY} t_y\}$$

$$= \exp\{t_x^T \mu_X + \frac{1}{2} t_x^T \Sigma_X t_x\} \cdot \exp\{t_y^T \mu_Y + \frac{1}{2} t_y^T \Sigma_Y t_y\}$$

Therefore, $\exp\{t_x^T \Sigma_{XY} t_y\} = 1 \Rightarrow t_x^T \Sigma_{XY} t_y = 0, \ \forall t_x, t_y$.

We can let t_x and t_y take the form $(0 \cdots 0 1 0 \cdots)^T$ to get the (i, j) entry of Σ_{XY}. By looping over (i, j), we can get $(\Sigma_{XY})_{ij} = 0. \ \forall i, j$.

Therefore, $\Sigma_{X,Y} = (\Sigma_{YX})^T = (0)$

2. Problem 2

(a) It's known that if $U_i, i = 1 \cdots n$ are iid, then the joint density of $U_{(1,n)} \cdots U_{(n,n)}$ is

$$f_{U_{(1,n)} \cdots U_{(n,n)}}(x_1 \cdots x_n) = \begin{cases} n! \, f_{U_1}(x_1) \cdots f_{U_n}(x_n), & -\infty < x_1 < \cdots < x_n < +\infty \\ 0, & \text{otherwise} \end{cases}$$

And then density of $f_{U_i}(x_i) = 1,\ 0<x_i<1$, so

$$f_{U_{(1,n)} \cdots U_{(n,n)}}(x_1 \cdots x_n) = \begin{cases} n!, & 0<x_1<\cdots<x_n<1 \\ 0, & \text{otherwise} \end{cases}$$

(b) It's known that $E_i \overset{iid}{\sim} Exp(\lambda),\ i=1 \cdots n+1$. so the joint distribution of $(E_1 \cdots E_{n+1})$ is

$$f(e_1 \cdots e_{n+1}) = \prod_{i=1}^{n+1} \lambda \exp(-\lambda e_i)$$
$$= \lambda^{n+1} \exp\left(-\lambda \sum_{i=1}^{n+1} e_i\right),\quad e_i \geq 0$$

Note that, the vector $(P_1 \cdots P_{n+1})$ can be derived by

$$\begin{bmatrix} P_1 \\ \vdots \\ P_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & \cdots & 1 \end{bmatrix} \begin{bmatrix} E_1 \\ \vdots \\ E_{n+1} \end{bmatrix},\ \text{i.e. } \vec{P} = A\vec{E}$$

Therefore, the joint density of $(P_1 \cdots P_{n+1})$ is

$$f(t_1 \cdots t_{n+1}) = |A|^{-1} f(e_1 \cdots e_{n+1})$$

we have $|A| = 1$, so

$$f(t_1 \cdots t_{n+1}) = \lambda^n \exp(-\lambda t_{n+1}),\quad 0<t_1<\cdots<t_{n+1}$$

Then the joint density of $\left(\frac{P_1}{P_{n+1}} \cdots \frac{P_n}{P_{n+1}}, P_{n+1}\right)$

$x_1 = \frac{t_1}{t_{n+1}}$

$$f(x_1 \cdots x_{n+1}) = |J|\, f(t_1 \cdots t_{n+1}),\ 0<x_1<\cdots<x_n<1,\ 0<x_{n+1}$$

where $$|J| = \begin{vmatrix} \frac{\partial t_1}{\partial x_1} & \cdots & \frac{\partial t_1}{\partial x_{n+1}} \\ \vdots & & \vdots \\ \frac{\partial t_{n+1}}{\partial x_1} & \cdots & \frac{\partial t_{n+1}}{\partial x_{n+1}} \end{vmatrix} = \begin{vmatrix} x_{n+1} & 0 & \cdots & x_1 \\ 0 & x_{n+1} & \cdots & x_2 \\ \vdots & & \ddots & \vdots \\ & & x_{n+1} & x_n \\ 0 & 0 & \cdots & 1 \end{vmatrix} = (x_{n+1})^n$$

So $f(x_1 \cdots x_{n+1}) = (x_{n+1})^n \cdot \lambda^{n+1} \exp(-\lambda x_{n+1})$,

$0 < x_1 < \cdots < x_n < 1$, $0 < x_{n+1}$

Finally, to get density of $f(x_1 \cdots x_n)$, we integrate over x_{n+1}.

that is

$$f(x_1 \cdots x_n) = \int_0^{+\infty} (x_{n+1})^n \lambda^{n+1} \exp(-\lambda x_{n+1})\, dx_{n+1}$$

$$= \int_0^{+\infty} (x_{n+1})^{n+1-1} \lambda^{n+1} \exp(-\lambda x_{n+1})\, dx_{n+1}$$

$$= \Gamma(n+1)$$

$$= n!, \quad 0 < x_1 < \cdots < x_n < 1$$

So. $(U_{(1,n)} \cdots U_{(n,n)}) \overset{d.}{=} \left(\frac{P_1}{P_{n+1}} \cdots \frac{P_n}{P_{n+1}}\right)$

(c) We consider the CDF of $\left(F^{-1}\left(\frac{P_1}{P_{n+1}}\right), \cdots F^{-1}\left(\frac{P_n}{P_{n+1}}\right)\right)$,

denote it by $H(v_1 \cdots v_n)$ then

$$H(v_1 \cdots v_n) = \mathbb{P}\left[F^{-1}\left(\frac{P_1}{P_{n+1}}\right) \leq v_1, \cdots F^{-1}\left(\frac{P_n}{P_{n+1}}\right) \leq v_n\right]$$

since F is strictly monotone, and F is CDF

$$H(v_1 \cdots v_n) = \mathbb{P}\left[\frac{P_1}{P_{n+1}} \leq F(v_1) \cdots \frac{P_n}{P_{n+1}} \leq F(v_n)\right]$$

since $(U_{(1,n)} \cdots U_{(n,n)}) \overset{d}{=} \left(\frac{P_1}{P_{n+1}} \cdots \frac{P_n}{P_{n+1}}\right)$, we have

$$H(v_1 \cdots v_n) = \mathbb{P}\left[U_{(1,n)} \leq F(v_1) \cdots U_{(n,n)} \leq F(v_n)\right]$$

$$= \mathbb{P}\left[F^{-1}(U_{(1,n)}) \leq v_1 \cdots F^{-1}(U_{(n,n)}) \leq v_n\right]$$

$$= \mathbb{P}\left[Y_{(1,n)} \leq v_1, \cdots Y_{(n,n)} \leq v_n\right]$$

The last step is because, for an iid sample $Y_1 \cdots Y_n$,

we can get them from iid sample $U_1 \cdots U_n$

of uniform $[0,1]$ by taking $Y_i = F^{-1}(U_i)$

This is an one-to-one map, so we have

$$(Y_{(1,n)} \cdots Y_{(n,n)}) \overset{d}{=} (F^{-1}(U_{(1,n)}) \cdots F^{-1}(U_{(n,n)}))$$

Therefore, we have

$$\mathbb{P}[F^{-1}(\frac{P_1}{P_{n+1}}) \le v_1, \cdots F^{-1}(\frac{P_n}{P_{n+1}}) \le v_n]$$

$$= \mathbb{P}[Y_{(1,n)} \le v_1, \cdots Y_{(n,n)} \le v_n], \ \forall v_i,$$

Hence

$$(Y_{(1,n)} \cdots Y_{(n,n)}) \overset{d}{=} (F^{-1}(\frac{P_1}{P_{n+1}}) \cdots F^{-1}(\frac{P_1}{P_{n+1}}))$$

3. Problem 3. — 5.24

From Thm 5.4.6, we have the density of $X_{(1)}, X_{(n)}$ is

$$f_{X_{(1)},X_{(n)}}(x_1, x_n) = \frac{n!}{(n-2)!} f_X(x_1) f_X(x_n) (F_X(x_n) - F_X(x_1))^{n-2}$$

It's known that

$$f_X(x) = \begin{cases} 1/\theta, & 0<x<\theta \\ 0, & \text{otherwise} \end{cases}$$

so cdf $F_X(x) = \frac{x}{\theta}$, $0<x<\theta$, then $\quad 0<x_1<x_n<\theta$

$$f_{X_{(1)},X_{(n)}}(x_1, x_n) = n(n-1)\cdot\frac{1}{\theta}\cdot\frac{1}{\theta}\cdot\left(\frac{x_n-x_1}{\theta}\right)^{n-2} = \frac{n(n-1)}{\theta^n}(x_n-x_1)^{n-2}$$

Therefore the density of $(\frac{X_{(1)}}{X_{(n)}}, X_{(n)})$ is

$$f(u,v) = \left|\begin{matrix} \frac{\partial x_1}{\partial u} & \frac{\partial x_1}{\partial v} \\ \frac{\partial x_n}{\partial u} & \frac{\partial x_n}{\partial v} \end{matrix}\right| f_{X_{(1)},X_{(n)}}(x_1, x_n)$$

$x_n = v$

$$= \left|\begin{matrix} v & u \\ 0 & 1 \end{matrix}\right| \frac{n(n-1)}{\theta^n}(v-uv)^{n-2}$$

$$= v \cdot \frac{n(n-1)}{\theta^n} v^{n-2} (1-u)^{n-2}$$

$$= \frac{n(n-1)}{\theta^n} v^{n-1} (1-u)^{n-2}$$

$$= g(v) \cdot h(u) , \quad 0<u<1, \ 0<v<\theta$$

Since $f(u,v)$ can be factorized, we have

$X_{(1)}/X_{(n)}$ and $X_{(n)}$ are independent.

3. Problem 3-5.25.

The joint distribution of $X_{(1)} \cdots X_{(n)}$ is.

$$f(x_1 \cdots x_n) = \frac{n!\, a^n}{\theta^{an}} x_1^{a-1} x_2^{a-1} \cdots x_n^{a-1}, \quad 0<x_1<\cdots<x_n<\theta$$

Consider the transform $Y_i = X_{(i)}/X_{(i+1)}$ $i=1, \cdots n-1$

$Y_n = X_{(n)}$, This is an one-to-one map and the

Jacobian is
$$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} = \begin{vmatrix} y_2 \cdots y_n & * & * & * \\ 0 & y_3 \cdots y_n & & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & & 1 \end{vmatrix}$$

$$= (y_2 \cdots y_n)\cdot(y_3 \cdots y_n) \cdots (y_{n-1} y_n) \cdot 1$$

$$= y_2 \cdot y_3^2\, y_4^3 \cdots y_{n-1}^{n-2}\, y_n^{n-1},$$

So

$$f(y_1 \cdots y_n) \qquad (y_1 \cdots y_n)^{a-1} \cdot y_2^{a}\, y_3^{2a} \cdots y_n^{(n-1)a}$$

$$= y_2 y_3^2 \cdots y_n^{n-1} \frac{n!\, a^n}{\theta^{an}} (y_1 \cdots y_n)^{a-1} (y_2 \cdots y_n)^{a-1} \cdots y_n^{a-1}$$

$$= \frac{n!\, a^n}{\theta^{an}} y_1^{a-1} y_2^{2a-1} \cdots y_n^{na-1},$$

$= f_1(y_1) f_2(y_2) \cdots f_n(y_n)$ $\quad 0 < y_i < 1, i = 1 \cdots n-1$

$0 < y_n < \theta$

Therefore, $f(y_1 \cdots y_n)$ can be factorized, so

$X_{(1)}/X_{(2)}, X_{(2)}/X_{(3)} \cdots X_{(n-1)}/X_{(n)}, X_{(n)}$ are

independent.

For the distribution of $X_{(i)}/X_{(i+1)}$, $i = 1 \cdots n-1$

we directly integrate the term y_i^{ia-1} to get

the normalization constant.

$$\int_0^1 y_i^{ai-1} dy_i = \frac{y_i}{ai}\Big|_0^1 = \frac{1}{ai},$$

Therefore. $f_{X_{(i)}/X_{(i+1)}}(y_i) = ia\, y_i^{ia-1}, \quad 0 < y_i < 1$

For $X_{(n)}$, we integrate the term y_n^{na-1}

$$\int_0^\theta y_n^{na-1} dy_n = \frac{y_n^{na}}{na}\Big|_0^\theta = \frac{\theta^{na}}{na}, \text{ so}$$

$$f_{X_{(n)}}(y_n) = \frac{na}{\theta^{na}} y_n^{na-1}, \quad 0 < y_n < \theta$$

4. Problem 4 — 5.30

By CLT. we have $\bar{X}_1 \overset{d}{\approx} N(\mu, \frac{\sigma^2}{n})$, $\bar{X}_2 \overset{d}{\approx} N(\mu, \frac{\sigma^2}{n})$.

Since $\bar{X}_1$ and $\bar{X}_2$ are from two independent samples, they are independent as well. So. $\bar{X}_1 - \bar{X}_2 \overset{d}{\approx} N(0, \frac{2\sigma^2}{n})$, denoting by X

Then $\mathbb{P}[|\bar{X}_1 - \bar{X}_2| < \frac{\sigma}{5}] = \mathbb{P}[|X| < \frac{\sigma}{5}]$

$= \mathbb{P}[\sqrt{\frac{n}{2\sigma^2}} |X| < \frac{1}{5}\sqrt{\frac{n}{2}}]$

$= \mathbb{P}[|N(0,1)| < \frac{1}{5}\sqrt{\frac{n}{2}}]$

$= \Phi\left(\frac{1}{5}\sqrt{\frac{n}{2}}\right) - \Phi\left(-\frac{1}{5}\sqrt{\frac{n}{2}}\right)$

$= 2\Phi\left(\frac{1}{5}\sqrt{\frac{n}{2}}\right) - 1 \approx 0.99$

so we need $\frac{1}{5}\sqrt{\frac{n}{2}} = 2.5758 \Rightarrow n = 331.73 \approx 332$

Therefore $n = 332$.

4. Problem 4 — 5.31

Since $\bar{X}$ is mean of a random sample of size 100, we have $\mathbb{E}\bar{X} = \mu$,

$var(\bar{X}) = \frac{1}{n} var(X) = \frac{\sigma^2}{n} = 0.09$

By chebychev's inequality,

$$\mathbb{P}[\,|\bar{X}-\mu| \geq k\sigma\,] \leq \frac{1}{k^2}, \text{ i.e.}$$

$\mathbb{P}[-k\sigma \leq \bar{X}-\mu \leq k\sigma] \geq 1 - \frac{1}{k^2}$. To make this value at least 0.90, we need $1 - \frac{1}{k^2} \geq 0.9 \Rightarrow k \geq \sqrt{10}$

Therefore the bound of $\bar{X}-\mu$ is at least $[-0.9487, 0.9487]$

By CLT, $\bar{X} \overset{d}{\approx} N(\mu, \frac{\sigma^2}{n}) = N(\mu, 0.09)$

$\rightarrow \bar{X}-\mu \overset{d}{\approx} N(0, 0.09)$. Let the bound be $\pm c$,

then $\mathbb{P}[-c \leq \bar{X}-\mu \leq c] \geq 0.9$

$\rightarrow \mathbb{P}\left[-c\sqrt{\frac{n}{\sigma^2}} \leq \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \leq c\sqrt{\frac{n}{\sigma^2}}\right] \geq 0.9$, and $\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$

so $2\Phi\left(c \cdot \sqrt{\frac{n}{\sigma^2}}\right) - 1 \geq 0.9 \Rightarrow c \geq 0.4935$

Therefore the bound of $\bar{X}-\mu$ is at least $[-0.4935, 0.4935]$

We can see that the Chebychev's bounds are nearly twice as wide as the CLT, which shows Chebychev's Inequality is conservative. Moreover, the sample size is 100, large enough to make CLT be a good approxiamation. So we can use CLT safely regardless of whether the true distribution of $\{X_i\}$ is normal or not.

5. Problem 5 – 5.32

(a) It's known that $X_n \xrightarrow{P} a$ so.

$$\forall \tilde{\varepsilon} > 0, \lim_{n\to\infty} \mathbb{P}[\,|X_n - a| < \tilde{\varepsilon}\,] = 1 \quad (*)$$

Note that $\mathbb{P}[\,|X_n-a| < \tilde{\varepsilon}\,] = \mathbb{P}[\,|\sqrt{X_n}-\sqrt{a}|\,|\sqrt{X_n}+\sqrt{a}| < \tilde{\varepsilon}\,]$

Then we have $\mathbb{P}[\,|Y_n-\sqrt{a}| < \varepsilon\,] = \mathbb{P}[\,|X_n-a|/|\sqrt{X_n}+\sqrt{a}| < \varepsilon\,]$

$$= \mathbb{P}[\,|X_n-a| < \varepsilon\cdot|\sqrt{X_n}+\sqrt{a}|$$

$$\geq \mathbb{P}[\,|X_n-a| < \varepsilon|\sqrt{a}|\,].$$

The last line is because we always have $\sqrt{a}+\sqrt{X_n} \geq \sqrt{a}$ and when we decrease the bound, the probability will decrease. Then, in the expression

$\forall \tilde{\varepsilon}\ \lim_{n\to\infty} \mathbb{P}[\,|X_n-a| < \tilde{\varepsilon}\,] = 1$, we take $\tilde{\varepsilon} = \varepsilon|\sqrt{a}|$

we have $\lim_{n\to\infty} \mathbb{P}[\,|Y_n-\sqrt{a}| < \varepsilon\,] \geq \lim_{n\to\infty} \mathbb{P}[\,|X_n-a| < \varepsilon|\sqrt{a}|\,] = 1$

However we must have $\lim_{n\to\infty} \mathbb{P}[\,|Y_n-\sqrt{a}| < \varepsilon\,] \leq 1$. So

$\forall \varepsilon > 0,\ \lim_{n\to\infty} \mathbb{P}[\,|Y_n-\sqrt{a}| < \varepsilon\,] = 1$, i.e. $Y_n \xrightarrow{P} \sqrt{a}$

For $Y_n' = a/X_n$, $\forall \varepsilon > 0$

$$\mathbb{P}[\,|Y_n' - 1| < \varepsilon\,] = \mathbb{P}\left[\,\left|\frac{a}{X_n} - 1\right| < \varepsilon\,\right]$$

$$= \mathbb{P}\left[\,1-\varepsilon < \frac{a}{X_n} < 1+\varepsilon\,\right]$$

① if $0 < \varepsilon < 1$,

$$\mathbb{P}[\,|Y_n'-1| < \varepsilon\,] = \mathbb{P}\left[\,\frac{a}{1+\varepsilon} < X_n < \frac{a}{1-\varepsilon}\,\right]$$

$= \mathbb{P}[a - \frac{a}{1+\varepsilon} < X_n < a + \frac{a}{1-\varepsilon}]$

$= \mathbb{P}[-\frac{a}{1+\varepsilon} < X_n - a < \frac{a}{1-\varepsilon}]$

$\geq \mathbb{P}[-\frac{a}{1+\varepsilon} < X_n - a < \frac{a}{1+\varepsilon}]$ since $\frac{a}{1+\varepsilon} \leq \frac{a}{1-\varepsilon}$

$= \mathbb{P}[|X_n - a| < \frac{a}{1+\varepsilon}]$

By taking $\tilde{\varepsilon} = \frac{a}{1+\varepsilon}$, we have

$\lim_{n\to\infty} \mathbb{P}[|Y_n' - 1| < \varepsilon] \geq \lim_{n\to\infty} \mathbb{P}[|X_n - a| < \frac{a}{1+\varepsilon}] = 1, \quad 0 < \varepsilon < 1$

② if $\varepsilon \geq 1$. note thate $\mathbb{P}[X_i > 0] = 1$. $\forall i$

$\mathbb{P}[|Y_n' - 1| < \varepsilon] = \mathbb{P}[1 - \varepsilon < \frac{a}{X_n} < 1 + \varepsilon]$

$= \mathbb{P}[0 < \frac{a}{X_n} < 1 + \varepsilon]$

$= \mathbb{P}[X_n > \frac{a}{1+\varepsilon}]$

$\geq \mathbb{P}[\frac{2a\varepsilon + a}{1+\varepsilon} > X_n > \frac{a}{1+\varepsilon}]$ [since we set an upper bound]

$= \mathbb{P}[\frac{a}{1+\varepsilon} - a < X_n - a < -\frac{a}{1+\varepsilon} + a]$

$= \mathbb{P}[|X_n - a| < a - \frac{a}{1+\varepsilon}]$

By taking $\tilde{\varepsilon} = a - \frac{a}{1+\varepsilon}$, we have $\lim_{n\to\infty} \mathbb{P}[|Y_n' - 1| < \varepsilon] \geq 1$

From ①②, $\forall \varepsilon$, $\lim_{n\to\infty} \mathbb{P}[|Y_n' - 1| < \varepsilon] \geq 1$. so.

$\lim_{n\to\infty} \mathbb{P}[|Y_n' - 1| < \varepsilon] = 1$, $\forall \varepsilon$. i.e. $Y_n' \xrightarrow{P} 1$

Hence. both $Y_n = \sqrt{X_n}$ and $Y_n' = \frac{a}{X_n}$ converge in probability.

(b) Example 5.5.3 define $S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$.

It's well known that, if the r.v X has finite mean μ, variance σ^2, skewness γ and kurtosis κ, then

$$\operatorname{var}(S_n^2) = \left(\kappa - \frac{n-3}{n-1}\right)\frac{\sigma^4}{n}$$

so we have $\lim_{n\to\infty} \operatorname{var}(S_n^2) = 0$. then by chebychev's inequality, we have $S_n^2 \xrightarrow{P} \sigma^2$

Therefore, from the "Y_n" part of (a), we have

$$\sqrt{S_n^2} = S_n \xrightarrow{P} \sigma .$$

from the "Y_n'" part of (a), we have $\frac{\sigma}{S_n} \xrightarrow{P} 1$

So, we prove $\sigma / S_n \xrightarrow{P} 1$.

5 Problem 5 — 5.33

It's known that $X_n \xrightarrow{d} X$, i.e. $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$, i.e.

$\lim_{n\to\infty} \mathbb{P}[X_n \leq x] = \mathbb{P}[X \leq x]$, $\forall x$. or

$\lim_{n\to\infty} \mathbb{P}[X_n > x] = \mathbb{P}[X > x]$.

From $\lim_{n\to\infty} \mathbb{P}[Y_n > c] = 1$, we have, $\forall |c| < \infty$

$\forall \tilde{\varepsilon} > 0$, $\exists N \in \mathbb{Z}$, s.t. $\forall n \geq N$, $\left| \mathbb{P}[Y_n > c] - 1 \right| < \tilde{\varepsilon} \Rightarrow \mathbb{P}[Y_n > c] > 1 - \tilde{\varepsilon}$

Now. $\forall \varepsilon > 0$, we choose $\tilde{\varepsilon} = \frac{\varepsilon}{2}$, then we have $\mathbb{P}[Y_n > c] > 1 - \frac{\varepsilon}{2}$.

Then, From $\lim_{n\to\infty} \mathbb{P}[X_n > x] = \mathbb{P}[X > x]$, we have

$\forall \delta > 0$, $\exists M \in \mathbb{Z}$, s.t. $\forall n \geq M$. $|\mathbb{P}[X_n > m] - \mathbb{P}(X > m)| < \delta$

i.e. $\mathbb{P}[X_n > x] > \mathbb{P}[X > x] - \delta$. we can choose a specific δ, s.t. $\delta < \frac{\varepsilon}{2}$, then we can choose m. s.t.

$\mathbb{P}[X > m] > 1 - \frac{\varepsilon}{2} + \delta$, then we have

$\exists m$, $\forall n \geq M$. $\mathbb{P}[X_n > x] > 1 - \frac{\varepsilon}{2} + \delta - \delta = 1 - \frac{\varepsilon}{2}$

Now, for this m, M, and N, $\forall n \geq \max\{N, M\}$ we have

$\mathbb{P}[X_n + Y_n > c] \geq \mathbb{P}[X_n > m, Y_n > c - m]$ (This is because we can always describe the event $X_n + Y_n > c$ using another different m)

$\geq \mathbb{P}[X_n > m] + \mathbb{P}[Y_n > c - m] - 1$ [by $P(AB) \geq P(A) + P(B) - 1$]

$> 1 - \frac{\varepsilon}{2} + 1 - \frac{\varepsilon}{2} - 1$

$= 1 - \varepsilon$

More clearly, we find a $\tilde{N} = \max\{N, M\}$. s.t.

$\forall \varepsilon > 0$. $\forall n \geq \tilde{N}$, $\mathbb{P}[X_n + Y_n > c] > 1 - \varepsilon$

i.e. $\left| \mathbb{P}[X_n + Y_n > c] - 1 \right| < \varepsilon$

Therefore $\lim_{n \to \infty} \mathbb{P}[X_n + Y_n > c] = 1$

6. Problem 6.

(a) By WLLN, we have $\bar{X}_n \xrightarrow{P} \mathbb{E}X = \frac{1}{\lambda}$.

Consider the function $g(x) = \frac{1}{x}$. Its only discontinuity point is $x = 0$, And obviously $\mathbb{P}[\bar{X}_n = 0] = 0$. So. by continuous mapping theorem, we have $g(\bar{X}_n) \xrightarrow{P} g(\frac{1}{\lambda})$, that is.

$\frac{1}{\bar{X}_n} \xrightarrow{P} \lambda$. So $\hat{\lambda}_n$ is a consistent estimator of λ.

(b) From CLT, we have

$$\sqrt{n}(\bar{X}_n - \frac{1}{\lambda}) \xrightarrow{d} N(0, \mathrm{var}(X)) = N(0, \frac{1}{\lambda^2})$$

Now consider the function $g(x) = \frac{1}{x}$. $g'(x) = -\frac{1}{x^2}$

By delta method, we have

$$\sqrt{n}(g(\bar{X}_n) - g(\frac{1}{\lambda})) \xrightarrow{d} N(0, \frac{1}{\lambda^2}[g'(\frac{1}{\lambda})]^2)$$

that is

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow{d} N(0, \frac{1}{\lambda^2} \cdot \lambda^4) = N(0, \lambda^2)$$

and we have asymptotic variance $\sigma^2 = \lambda^2$