MATH50010: Probability for Statistics
Problem Sheet 4

1. The joint pdf of the random variables X_1 and X_2 is

$$f_{X_1,X_2}(x_1,x_2) = k\exp\left\{-\left(\frac{x_1^2}{6} - \frac{x_1x_2}{3} + \frac{2x_2^2}{3}\right)\right\}, \text{ for } -\infty < x_1, x_2 < \infty.$$

Find $\mathrm{E}(X_1), \mathrm{E}(X_2), \mathrm{Var}(X_1), \mathrm{Var}(X_2), \mathrm{Cov}(X_1, X_2)$ and k.

Because the log pdf is quadratic in x_1 and x_2, and pdfs must integrate to 1, we see that (X_1, X_2) must follow a bivariate normal distribution. Further since the exponent lacks a constant term, $\mu_1 = \mathrm{E}(X_1)$ and $\mu_2 = \mathrm{E}(X_2)$ are both zero. Thus the joint pdf of (X_1, X_2) is

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{y_1^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2} - \frac{2\rho y_1y_2}{\sigma_1\sigma_2}\right)\right\},$$

where

$$2(1-\rho^2)\sigma_1^2 = 6, \quad 2(1-\rho^2)\sigma_2^2 = \frac{3}{2}, \text{ and } \frac{\sigma_1\sigma_2(1-\rho^2)}{\rho} = 3.$$

Solving for σ_1^2 and σ_2^2 in first two equations and substituting into the square of the third equation gives $\frac{9}{4\rho^2} = 9$, and thus $\rho = 1/2, \sigma_1^2 = 4$, and $\sigma_2^2 = 1$. Finally, by the properties of the bivariate normal distribution, $\mathrm{Var}(X_1) = 4, \mathrm{Var}(X_2) = 1, \mathrm{Cov}(X_1, X_2) = \rho\sigma_1\sigma_2 = 1$, and $k = (2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1} = 1/(2\sqrt{3}\pi)$.

2. Suppose

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\left[\mu = \begin{pmatrix} 2 \\ -5 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 4 \end{pmatrix}\right].$$

Compute $\Pr(X_1 > 0)$ and $\Pr(X_2 < -6)$.

Using the results in the lecture notes, $X_1 \sim \mathrm{N}(2, 1)$ and $X_2 \sim \mathrm{N}(-5, 4)$. Thus

$$\Pr(X_1 > 0) = \Pr(Z > -2) = 1 - \Phi(-2) = \Phi(2) = 97.73\%,$$

where Z is a standard normal random variable. Likewise,

$$\Pr(X_2 < -6) = \Pr\left(Z < \frac{-6+5}{2}\right) = \Phi(-1/2) = 1 - \Phi(1/2) = 30.85\%.$$

3. Suppose X_1, X_2, and X_3 are iid $\mathrm{N}(1, 1)$ random variables. Let $X_4 = 2X_2 + 2X_3$ and $X_5 = X_2 - 2X_3$.

   (a) Find the joint pdf of (X_1, X_4, X_5).

   (b) Find the marginal pdf of X_5.

   (a) Let $\boldsymbol{U} = (X_2, X_3)^\top$, $\boldsymbol{V} = (X_4, X_5)^\top$, and $\boldsymbol{M} = \begin{pmatrix} 2 & 2 \\ 1 & -2 \end{pmatrix}$, so that $\boldsymbol{V} = \boldsymbol{M}\boldsymbol{U}$. Because X_2 and X_3 are independent normal random variables, they are jointly bivariate normal. (This can be verified by comparing their joint pdf with that of a bivariate normal random variable.) Thus

$$\boldsymbol{U} \sim \mathrm{N}_2\left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right)$$

and

$$\boldsymbol{V} \sim \mathrm{N}_2\left(\boldsymbol{M}\begin{pmatrix}1\\1\end{pmatrix}, \boldsymbol{M}\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\boldsymbol{M}^\top\right),$$

i.e.,

$$\boldsymbol{V} \sim \mathrm{N}_2\left(\begin{pmatrix}4\\-1\end{pmatrix}, \begin{pmatrix}8 & -2\\-2 & 5\end{pmatrix}\right). \tag{1}$$

Finally, because X_1 and $\boldsymbol{V}$ are independent normal random vectors, they are jointly multivariate normal, so

$$\begin{pmatrix}X_1\\X_4\\X_5\end{pmatrix} \sim \mathrm{N}_3\left(\begin{pmatrix}1\\4\\-1\end{pmatrix}, \begin{pmatrix}1 & 0 & 0\\0 & 8 & -2\\0 & -2 & 5\end{pmatrix}\right).$$

(b) *By (**??**) X_4 and X_5 are bivariate normal, so by results derived in lecture notes, $X_5 \sim \mathrm{N}(-1, 5)$.*

4. Suppose that X and Y are absolutely continuous random variables with pdf given by

$$f_{X,Y}(x,y) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}\left(x^2+y^2\right)\right\}, \quad \text{for } x, y \in \mathbb{R}.$$

(a) Let the random variable U be defined by $U = X/Y$. Find the pdf of U. Do you recognize the distribution of U?

(b) Suppose now that $S \sim \chi^2_\nu$ is independent of X and Y. (The pdf of S is given by

$$f_S(s) = c(\nu)s^{\nu/2-1}e^{-s/2}, \quad \text{for } s > 0,$$

where ν is a positive integer and $c(\nu)$ is a normalizing constant depending on ν.) Find the pdf of random variable T defined by

$$T = \frac{X}{\sqrt{S/\nu}}.$$

Show that this is the pdf of a t random variable with ν degrees of freedom.

(a) *Put $U = X/Y$ and $V = Y$; the inverse transformations are therefore $X = UV$ and $Y = V$. In terms of the multivariate transformation theorem, we have transformation functions defined by*

$$g_1(x,y) = x/y, \qquad g_1^{-1}(u,v) = uv,$$

$$g_2(x,y) = y, \qquad g_2^{-1}(u,v) = v,$$

and the Jacobian of the transformation is given by

$$J(u,v) = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$$

and hence

$$f_{U,V}(u,v) = f_{X,Y}(uv,v)\ |v| = \left(\frac{1}{2\pi}\right)\exp\left\{-\frac{1}{2}(u^2v^2+v^2)\right\}|v|\,,\ \textit{for}\ (u,v)\in\mathbb{R}^2.$$

Now, for any real u*,*

$$\begin{aligned} f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u,v)\,\mathrm{d}v &= \int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\right)\exp\left\{-\frac{1}{2}(u^2v^2+v^2)\right\}|v|\ \mathrm{d}v \\ &= \left(\frac{1}{\pi}\right)\int_0^{\infty} v\exp\left\{-\frac{v^2}{2}(1+u^2)\right\}\,\mathrm{d}v \quad \textit{(as integrand is even function)} \\ &= \left(\frac{1}{\pi}\right)\left[-\frac{1}{(1+u^2)}\exp\left\{-\frac{v^2}{2}(1+u^2)\right\}\right]\Bigg|_0^{\infty} = \frac{1}{\pi(1+u^2)}, \end{aligned}$$

with the final step following by direct integration.

Reflect: This is the infamous Cauchy distribution. It is most commonly encountered as a case where results such as the central limit theorem do not apply, because U does not have finite moments of any order. In particular, it does not have a finite mean or variance.

As its pdf is an even function, we might be tempted to assert that its mean should be zero. But this does not respect the definition we made when defining expectation. $E|U|$ is not finite, so $E(U)$ is not defined. Equivalently, its mean is zero only if we allow $\infty - \infty = 0$, which we do not.

(b) Now put $T = X/\sqrt{S/\nu}$ *and* $R = S$*; the inverse transformations are therefore* $X = T\sqrt{R/\nu}$ *and* $S = R$*. In terms of the multivariate transformation theorem, we have transformation functions from* $(X,S) \to (T,R)$ *defined by*

$$g_1(x,s) = x/\sqrt{s/\nu}, \qquad g_1^{-1}(t,r) = t\sqrt{r/\nu},$$

$$g_2(x,s) = s, \qquad g_2^{-1}(t,r) = r,$$

and the Jacobian of the transformation is given by

$$J(t,r) = \begin{vmatrix} \sqrt{\dfrac{r}{\nu}} & \dfrac{t}{2\sqrt{r\nu}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{r}{\nu}},$$

and hence

$$f_{T,R}(t,r) = f_{X,S}\left(t\sqrt{\frac{r}{\nu}}, r\right)\sqrt{\frac{r}{\nu}} \quad = \quad f_X\left(t\sqrt{\frac{r}{\nu}}\right)\, f_S(r)\sqrt{\frac{r}{\nu}},\ \textit{for}\ t\in\mathbb{R}, r\in\mathbb{R}^+,$$

and zero otherwise. Now, for any real t*,*

$$
\begin{aligned}
f_T(t) &= \int_{-\infty}^{\infty} f_{T,R}(t,r)\,\mathrm{d}r \\
&= \int_0^{\infty} \left(\frac{1}{2\pi}\right)^{1/2} \exp\left\{-\frac{rt^2}{2\nu}\right\} c(\nu) r^{\nu/2-1} e^{-r/2} \sqrt{\frac{r}{\nu}}\,\mathrm{d}r \\
&= \left(\frac{1}{2\pi}\right)^{1/2} \frac{c(\nu)}{\sqrt{\nu}} \int_0^{\infty} r^{(\nu+1)/2-1} \exp\left\{-\frac{r}{2}\left(1+\frac{t^2}{\nu}\right)\right\}\,\mathrm{d}r \\
&= \left(\frac{1}{2\pi}\right)^{1/2} \frac{c(\nu)}{\sqrt{\nu}} \left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2} \int_0^{\infty} z^{(\nu+1)/2-1} \exp\left\{-\frac{z}{2}\right\}\,\mathrm{d}z \quad \textit{setting } z = r\left(1+\frac{t^2}{\nu}\right) \\
&= \left(\frac{1}{2\pi}\right)^{1/2} \frac{c(\nu)}{\sqrt{\nu}} \left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2} \frac{1}{c(\nu+1)},
\end{aligned}
$$

as the integrand is proportional to a Gamma pdf. We also can see that f_S is a $\mathrm{Gamma}(\nu/2, 1/2)$ (otherwise known as a $\chi^2(\nu)$ or χ^2_ν) density, and that the normalizing constant $c(\nu)$ is given by

$$
c(\nu) = \frac{\left(\frac{1}{2}\right)^{\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)} \qquad \Longrightarrow \qquad f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{1}{\pi\nu}\right)^{1/2} \frac{1}{(1+t^2/\nu)^{(\nu+1)/2}},
$$

which, in fact, is the density of the $\mathrm{Student}(\nu)$ or t_ν distribution. You will recall this from your work on the t-test last year.

5. Suppose that U_1 and U_2 are independent and identically distributed $\mathrm{Unif}(0,1)$ random variables. Let random variables Z_1 and Z_2 be defined by

$$
\begin{aligned}
Z_1 &= \sqrt{-2\log(U_1)}\cos(2\pi U_2), \\
Z_2 &= \sqrt{-2\log(U_1)}\sin(2\pi U_2).
\end{aligned}
$$

Find the joint pdf of (Z_1, Z_2).

The inverse of the transformation

$$
\begin{cases}
Z_1 &= \sqrt{-2\log(U_1)}\cos(2\pi U_2) \\
Z_2 &= \sqrt{-2\log(U_1)}\sin(2\pi U_2)
\end{cases}
$$

is

$$
\begin{cases}
U_1 &= \exp\left\{-\frac{1}{2}\left(Z_1^2+Z_2^2\right)\right\} \\
U_2 &= I\{Z_2>0\}\Big(\frac{1}{2\pi}\arccos\frac{Z_1}{\sqrt{Z_1^2+Z_2^2}}\Big) + I\{Z_2<0\}\Big(1-\frac{1}{2\pi}\arccos\frac{Z_1}{\sqrt{Z_1^2+Z_2^2}}\Big)
\end{cases},
$$

where $I\{\}$ is an indicator function and the $\arccos$ *function has a range of $(0, \pi)$. Notice, from the definition of Z_2, that $Z_2 < 0$ if and only if $U_2 > \frac{1}{2}$. The range of the new variables is $\mathbb{R} \times \mathbb{R}$. When $Z_2 > 0$, the Jacobian of the transformation is*

$$
\begin{aligned}
\left| \begin{array}{cc} \frac{\partial u_1}{\partial z_1} & \frac{\partial u_1}{\partial z_2} \\ \frac{\partial u_2}{\partial z_1} & \frac{\partial u_2}{\partial z_2} \end{array} \right| &= \left| \begin{array}{cc} -z_1 \exp\left\{-\frac{1}{2}\left(z_1^2+z_2^2\right)\right\} & -z_2 \exp\left\{-\frac{1}{2}\left(z_1^2+z_2^2\right)\right\} \\ -\frac{1}{2\pi}\frac{z_2}{z_1^2+z_2^2} & \frac{1}{2\pi}\frac{z_1}{z_1^2+z_2^2} \end{array} \right| \\
&= \left| \frac{1}{2\pi}\frac{z_1^2}{z_1^2+z_2^2} \exp\left\{-\frac{1}{2}\left(z_1^2+z_2^2\right)\right\} + \frac{1}{2\pi}\frac{z_2^2}{z_1^2+z_2^2} \exp\left\{-\frac{1}{2}\left(z_1^2+z_2^2\right)\right\} \right| \\
&= \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left(z_1^2+z_2^2\right)\right\}.
\end{aligned}
$$

When $Z_2 < 0$, the signs on $\partial u_2 / \partial z_1$ and $\partial u_2 / \partial z_2$ change, but this does not affect the absolute value of the Jacobian. Hence the joint pdf is

$$
\begin{aligned}
f_{Z_1,Z_2}(z_1, z_2) &= f_{U_1,U_2}\left(\exp\left\{-\frac{1}{2}\left(z_1^2+z_2^2\right)\right\}, \frac{1}{2\pi}\arctan\frac{z_2}{z_1}\right) J(z_1, z_2) \\
&= 1 \times \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left(z_1^2+z_2^2\right)\right\} = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left(z_1^2+z_2^2\right)\right\},
\end{aligned}
$$

for $(z_1, z_2) \in \mathbb{R}^2$. Note that

$$
f_{Z_1,Z_2}(z_1, z_2) = f_{Z_1}(z_1) f_{Z_2}(z_2),
$$

where

$$
f_{Z_1}(z_1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z_1^2\right\}, \qquad f_{Z_2}(z_2) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z_2^2\right\},
$$

so in fact Z_1 and Z_2 are independent standard Normal random variables.

Reflect: This result tells us that if we can easily simulate (pairs of independent) uniform random variables, then we can easily simulate (pairs of independent) standard normal variables. This is extremely useful in practice! The result here is called the Box-Muller transform, and it is reviewed in an R tutorial on Blackboard. It is intimately connected to the polar coordinates trick for evaluating the Gaussian integral.

6. Suppose $(X_1, \ldots, X_n)$ is a collection of independent and identically distributed random variables taking values on $\mathbb{X}$ with pmf/pdf f_X and cdf F_X. Let Y_n and Z_n correspond to the *maximum* and *minimum* order statistics derived from $(X_1, \ldots, X_n)$, that is

$$
Y_n = \max\{X_1, \ldots, X_n\}, \quad Z_n = \min\{X_1, \ldots, X_n\}.
$$

(a) Show that the cdfs of Y_n and Z_n are given by

$$
F_{Y_n}(y) = \{F_X(y)\}^n, \quad F_{Z_n}(z) = 1 - \{1 - F_X(z)\}^n.
$$

(b) Suppose $X_1, \ldots, X_n \sim \text{Unif}(0, 1)$, that is

$$
F_X(x) = x, \text{ for } 0 \leq x \leq 1.
$$

Find the cdfs of Y_n and Z_n.

(c) Suppose $X_1, \ldots, X_n$ have cdf

$$F_X(x) = 1 - x^{-1}, \text{ for } x \geq 1.$$

Find the cdfs of Z_n and $U_n = Z_n^n$.

(d) Suppose $X_1, \ldots, X_n$ have cdf

$$F_X(x) = \frac{1}{1+e^{-x}}, \text{ for } x \in \mathbb{R}.$$

Find the cdfs of Y_n and $U_n = Y_n - \log n$.

(e) Suppose $X_1, \ldots, X_n$ have cdf

$$F_X(x) = 1 - \frac{1}{1+\lambda x}, \text{ for } x > 0.$$

Find the cdfs of Y_n, Z_n, $U_n = Y_n/n$, and $V_n = nZ_n$.

(a) From first principles,

$$\begin{aligned} F_{Y_n}(y) &= \Pr(Y_n \leq y) = \Pr\Big(\max\{X_1, \ldots, X_n\} \leq y\Big) = \Pr\Big(X_1 \leq y, \ldots, X_n \leq y\Big) \\ &= \prod_{i=1}^n \Pr(X_i \leq y) = \prod_{i=1}^n F_X(y) = \{F_X(y)\}^n . \end{aligned}$$

Likewise,

$$\begin{aligned} \Pr(Z_n > z) &= \Pr\Big(\min\{X_1, \ldots, X_n\} > z\Big) = \Pr\Big(X_1 > z, \ldots, X_n > z\Big) \\ &= \prod_{i=1}^n \Pr(X_i > z) = \prod_{i=1}^n \{1 - F_X(z)\} = \{1 - F_X(z)\}^n . \end{aligned}$$

So that $F_{Z_n}(z) = 1 - \Pr(Z_n > z) = 1 - \{1 - F_X(z)\}^n$.

(b) Directly applying the formulae derived in part (a),

$$F_{Y_n}(y) = \{F_X(y)\}^n = y^n$$

and

$$F_{Z_n}(z) = 1 - \{1 - F_X(z)\}^n = 1 - (1-z)^n.$$

(c) Again,

$$F_{Z_n}(z) = 1 - \{1 - F_X(z)\}^n = 1 - \left(1 - \left(1 - \frac{1}{z}\right)\right)^n = 1 - \frac{1}{z^n}, \qquad z \geq 1.$$

Setting $U_n = Z_n^n$*, we have from first principles that, for* $u > 1$*,*

$$F_{U_n}(u) = \Pr\left(U_n \leq u\right) = \Pr\left(Z_n^n \leq u\right) = \Pr\left(Z_n \leq u^{1/n}\right) = 1 - \frac{1}{\left(u^{1/n}\right)^n} = 1 - \frac{1}{u},$$

which is a valid cdf, but which does not depend on n*. Hence the limiting distribution of* U_n *is precisely*

$$F_U(u) = 1 - \frac{1}{u}, \quad u > 1.$$

(d) And again,

$$F_{Y_n}(y) = \{F_X(y)\}^n = \left(\frac{1}{1+e^{-y}}\right)^n, \qquad y \in \mathbb{R}.$$

Setting $U_n = Y_n - \log n$, we have from first principles that,

$$\begin{aligned} F_{U_n}(u) &= \Pr(U_n \le u) = \Pr(Y_n - \log n \le u) \\ &= \Pr(Y_n \le u + \log n) = F_{Y_n}(u + \log n) = \left(\frac{1}{1+e^{-u-\log n}}\right)^n. \end{aligned}$$

(e) And once again applying the formula from (a),

$$F_{Y_n}(y) = \{F_X(y)\}^n = \left(\frac{\lambda y}{1+\lambda y}\right)^n, \text{ for } y > 0,$$

and

$$F_{Z_n}(z) = 1 - \{1 - F_X(z)\}^n = 1 - \left(1 - \left(1 - \frac{1}{1+\lambda z}\right)\right)^n = 1 - \frac{1}{(1+\lambda z)^n}.$$

Now, setting $U_n = Y_n/n$, we have from first principles that, for $u > 0$,

$$F_{U_n}(u) = \Pr(U_n \le u) = \Pr(Y_n/n \le u) = \Pr(Y_n \le nu) = F_{Y_n}(nu) = \left(\frac{\lambda nu}{1+\lambda nu}\right)^n.$$

And setting $V_n = nZ_n$, we have from first principles that, for $v > 0$,

$$F_{V_n}(v) = \Pr(V_n \le v) = \Pr(nZ_n \le v) = \Pr(Z_n \le v/n) = F_{Z_n}(v/n) = 1 - \left(\frac{1}{1+\frac{\lambda v}{n}}\right)^n.$$

For discussion

7. Let $X_1, \dots X_n \sim \textsc{Uniform}(0,1)$ and let $M_n = \max\{X_1, \dots X_n\}$.

   (a) Show that for $\epsilon > 0$,
   $$\Pr(M_n < 1-\epsilon) = (1-\epsilon)^n.$$

   (b) Use the result above to show that for all $\epsilon > 0$
   $$\lim_{n\to\infty} \Pr(|M_n - 1| \geq \epsilon) = 0.$$
   Later we will say that this shows that the random variable M_n converges in probability to the constant value 1.

   (c) Now (by taking $\epsilon = \frac{t}{n}$), show that the distribution function of the rescaled variable $n(1-M_n)$ converges to the CDF of a known distribution.

   (a) *$M_n < x$ iff $X_i < x$ for each $i = 1, \dots, n$ so by independence:*
   $$\Pr(M_n < 1-\epsilon) = \prod_{i=1}^{n} \Pr(X_i < 1-\epsilon) = (1-\epsilon)^n.$$

   (b)
   $$\Pr(|M_n - 1| \geq \epsilon) = \Pr(M_n > 1+\epsilon) + \Pr(M_n \leq 1-\epsilon) = 0 + (1-\epsilon)^n \to 0$$

   (c)
   $$\Pr\left(n\left(1-M_n\right) \leq t\right) = \Pr\left(M_n \geq 1 - \frac{t}{n}\right) = 1 - \Pr\left(M_n < 1 - \frac{t}{n}\right) = 1 - \left(1 - \frac{t}{n}\right)^n \to 1 - \exp(-t).$$
   This is the CDF of a rate 1 exponential variable.

8. Suppose Y and $\boldsymbol{X} = (X_1, X_2)^\top$ jointly follow a trivariate normal distribution. Here Y is a univariate random variable and $\boldsymbol{Z} = (Y, X_1, X_2)^\top$ is a (3×1) trivariate normal random vector with mean
   $$\boldsymbol{\mu} = \begin{pmatrix} \mu_Y \\ \boldsymbol{\mu_X} \end{pmatrix} \text{ and variance-covariance matrix } \boldsymbol{M}^{-1} = \begin{pmatrix} m_{YY} & \boldsymbol{M}_{Y\boldsymbol{X}} \\ \boldsymbol{M}_{Y\boldsymbol{X}}^\top & \boldsymbol{M}_{\boldsymbol{XX}} \end{pmatrix}^{-1},$$
   where μ_Y is the univariate mean of Y, $\boldsymbol{\mu_X}$ is the (2×1) mean vector of $\boldsymbol{X}$, $\boldsymbol{\mu}$ is the (3×1) mean vector of both $\boldsymbol{X}$ and Y, m_{YY} is the first diagonal element of $\boldsymbol{M}$, $\boldsymbol{M}_{\boldsymbol{XX}}$ is the lower-right (2×2) submatrix of $\boldsymbol{M}$, and $\boldsymbol{M}_{Y\boldsymbol{X}}$ is the remaining off-diagonal (1×2) submartix of $\boldsymbol{M}$. (Note that we parameterize the multivariate normal in terms of the *inverse* of its variance-covariance matrix. This will significantly simplify calculations!)

   (a) Derive the conditional distribution of Y given both X_1 and X_2. *[Hint: Use vector/matrix notation.]*

   (b) Now suppose Y and $\boldsymbol{X} = (X_1, \dots, X_n)^\top$ jointly follow a multivariate normal distribution. Here Y remains a univariate random variable and $\boldsymbol{Z} = (Y, X_1, \dots, X_n)^\top$ is an $[(n+1) \times 1]$ multivariate normal random vector. Use the same notation for the mean and the inverse of the variance-covariance matrix, but with appropriately adjusted dimensions. Derive the conditional distribution of Y given $X_1, \dots, X_n$. *[Hint: If you used vector/matrix notation in part (a), this problem will be very easy. If you did not, it will be very hard!]*

(c) Set $n = 1$ and check that your answer is the same as the conditional distribution for the bivariate normal derived in lecture.

(a) The conditional distribution of Y given $\boldsymbol{X} = (X_1, X_2)$ is proportional to the joint distribution of (Y, X_1, X_2), using $\boldsymbol{z}$ as short hand for $(y, x_1, x_2)^\top$,

$$\begin{aligned}
f_{Y|X_1,X_2}(y|x_1,x_2) &\propto f_{Y,X_1,X_2}(y,x_1,x_2) = |\boldsymbol{M}|^{1/2}(2\pi)^{-3/2}\exp\left\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})^\top \boldsymbol{M}(\boldsymbol{z}-\boldsymbol{\mu})\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})^\top \boldsymbol{M}(\boldsymbol{z}-\boldsymbol{\mu})\right\} \\
&\propto \exp\left\{-\frac{1}{2}\begin{pmatrix} y-\mu_Y \\ \boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{X}}\end{pmatrix}^\top \begin{pmatrix} m_{YY} & \boldsymbol{M}_{Y\boldsymbol{X}} \\ \boldsymbol{M}_{Y\boldsymbol{X}}^\top & \boldsymbol{M}_{\boldsymbol{X}\boldsymbol{X}}\end{pmatrix}\begin{pmatrix} y-\mu_Y \\ \boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{X}}\end{pmatrix}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\Big[(y-\mu_Y)^2 m_{YY} + 2(y-\mu_Y)\boldsymbol{M}_{Y\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{X}})\Big]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[y^2 m_{YY} - 2y\, m_{YY}\left(\mu_Y - \frac{\boldsymbol{M}_{Y\boldsymbol{X}}}{m_{YY}}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{X}})\right)\right]\right\}.
\end{aligned}$$

In the first three lines we simply omit the normalizing constant for the trivariate normal and rewrite the pdf in our specialized notation. In the fourth line we expand the matrix product, dropping the $\boldsymbol{M}_{\boldsymbol{X}\boldsymbol{X}}$ term because it does not involve y. In the last line we expand $(y-\mu_Y)^2$, collect terms that are quadratic and linear in y, and drop remaining terms that do not involve y. Next, we pull out m_{YY} and complete the square (in y) by adding a constant term, $f_{Y|X_1,X_2}(y|x_1,x_1)$

$$\begin{aligned}
&\propto \exp\left\{-\frac{m_{YY}}{2}\left[y^2 - 2y\left(\mu_Y - \frac{\boldsymbol{M}_{Y\boldsymbol{X}}}{m_{YY}}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{X}})\right) + \left(\mu_Y - \frac{\boldsymbol{M}_{Y\boldsymbol{X}}}{m_{YY}}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{X}})\right)^2\right]\right\} \\
&\propto \exp\left\{-\frac{m_{YY}}{2}\left[y - \left(\mu_Y - \frac{\boldsymbol{M}_{Y\boldsymbol{X}}}{m_{YY}}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{X}})\right)\right]^2\right\}.
\end{aligned}$$

Since the log of the pdf is quadratic in y, we find

$$Y|\boldsymbol{X} \sim \mathrm{N}\left(\mu_Y - \frac{\boldsymbol{M}_{Y\boldsymbol{X}}}{m_{YY}}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{X}}),\ \frac{1}{m_{YY}}\right). \tag{2}$$

Notice that we employ our general strategy for computing conditional distributions. Start with the joint distribution, remove all the constant factors, and identify the remaining distribution and its normalizing constant.

*(b) We use the same notation as in part (a), noting that $\boldsymbol{\mu}_{\boldsymbol{X}}$ is now $(n\times 1)$, $\boldsymbol{M}_{Y\boldsymbol{X}}$ is now $(1\times n)$, and $\boldsymbol{M}_{\boldsymbol{X}\boldsymbol{X}}$ is $(n\times n)$. With these changes the calculations are exactly the same as in part (a). (Well almost exactly the same, the power on 2π should be $-(n+1)/2$ instead of $-3/2$ but that does not change anything.) The conditional distribution of Y given $\boldsymbol{X}$ is exactly the same as that given in (**??**).*

(c) We can derive $\boldsymbol{M}$ in the bivariate case by inverting the variance-covariance matrix,

$$\Sigma^{-1} = \begin{pmatrix} \sigma_Y^2 & \sigma_Y\sigma_X\rho \\ \sigma_Y\sigma_X\rho & \sigma_X^2\end{pmatrix}^{-1} = \frac{1}{\sigma_Y^2\sigma_X^2(1-\rho^2)}\begin{pmatrix} \sigma_X^2 & -\sigma_Y\sigma_X\rho \\ -\sigma_Y\sigma_X\rho & \sigma_Y^2\end{pmatrix} = \boldsymbol{M}.$$

*Substituting $M_{YX} = -\dfrac{\rho}{\sigma_Y\sigma_X(1-\rho^2)}$ and $m_{YY} = \dfrac{1}{\sigma_Y^2(1-\rho^2)}$ into (**??**), gives*

$$Y|X \sim \mathrm{N}\left(\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x-\mu_X),\ \sigma_Y^2(1-\rho^2)\right).$$