Bryant Willoughby

Stats 510 Final Exam (LN), December 13, 2024

Name (first, last): | **UM ID** :

Please fill the boxes with your name/id and return this problem set stapled along with your solution papers.

1. (16pt) Alice rolled the dice independently n times, for some $n \in \mathbb{N}$, and recorded the outcome of her experiment using $X_i \in \{1, 6\}$, for $i = 1, \ldots, n$. She wanted to keep track of certain statistics by setting $Y_i = 0$ if $X_i \in \{3, 6\}$, $Y_i = 1$ if $X_i \in \{1, 4\}$, and $Y_i = -1$ if $X_i \in \{2, 5\}$.
   (i) (6pt) Provide the pmf of $Y_1 + Y_2$ and $Y_1 Y_2$.
   (ii) (6pt) For general n, find the expectation and variance of $Y_1 + \ldots + Y_n$, and an approximation expression for its cumulative distribution function. Justify your approximation.
   (iii) For $c > 0$, show $P(|Y_1 \ldots Y_n| \geq c) \leq C$, where $C \downarrow 0$ as $n \to \infty$ (Hint: use Chebychev's inequality).

2. (10pt) Let U and V be i.i.d. random variables and distributed according to the uniform distribution on $[0, 1]$.
   (i) (4pt) Derive an expression for the pdf of random variables $X := U + V$ and $Y := U^2$.
   (ii) (3pt) Prove that X and Y are *not* independent random variables.
   (iii) (3pt) Determine the joint pdf for the bivariate random vector (X, Y).

3. (6pt) Which of the following classes of distributions belong to the exponential families (justify!)
   (i) discrete distributions on integers;
   (ii) the negative binomial distributions, $P(X = x|r, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$, $x = r, r+1, \ldots$;
   (iii) mixture distributions of two or more Gamma distributions, i.e., those with the pdf of the form $f(x|k, p, \alpha, \beta) = \sum_{i=1}^{k} p_i \text{Gamma}(x|\alpha_i, \beta_i)$, for some $k \in \mathbb{N}$, $\alpha_i, \beta_i > 0$ for $i = 1, \ldots, k$.

4. (8pt) Suppose that a signal of interest is represented by a random variable X and we assume that $X \sim \text{Normal}(\mu, \sigma^2)$. The signal is contaminated by an amount of noise ϵ, so what is observed is $Y = X + \epsilon$. The noise ϵ is not independent of X; in fact, $\epsilon = aX + Z$, where a is a constant, Z is independent of X and $Z \sim \text{Normal}(0, 1)$.
   (i) (4pt) Argue that (X, Y) is a bivariate normal vector. What is its mean and covariance matrix?
   (ii) (2pt) Suppose that $corr(X, Y) < 0$. What does that say about the constant a?
   (ii) (2pt) Given an observation $Y = y$, derive $\mathbb{E}[X|Y = y]$.

5. (10pt) Let p, Z, X be random variables whose joint distribution is given by the hierarchical model:

$$
\begin{aligned}
p &\sim \text{Beta}(\alpha, \beta), && (1)\\
Z|p &\sim \text{Bernoulli}(p) && (2)\\
X|p, Z = 0 &\sim \text{Normal}(\mu_0, 1) && (3)\\
X|p, Z = 1 &\sim \text{Normal}(\mu_1, 1), && (4)
\end{aligned}
$$

   where parameters $\alpha, \beta > 0$, $\mu_0, \mu_1 \in \mathbb{R}$.
   (i) (5pt) Derive the pmf/pdf for the (marginal) distribution for the variable Z, and the conditional distribution of X given Z (that is, the conditional pdf of X given $Z = z$ for $z \in \{0, 1\}$).
   (ii) (3pt) Find the (marginal) expectation of X. (Hint: use iterated expectation formula).
   (iii) (2pt) Derive the conditional distribution for Z given X (that is, find the expression for $P(Z = z|X = x)$ for $z \in \{0, 1\}$).

$\{1, 2, \ldots, 6\}$ (ch 3(4))

1. (16pt) Alice rolled the dice independently n times, for some $n \in \mathbb{N}$, and recorded the outcome of her experiment using $X_i \in \{1, 6\}$, for $i = 1, \ldots, n$. She wanted to keep track of certain statistics by setting $Y_i = 0$ if $X_i \in \{3, 6\}$, $Y_i = 1$ if $X_i \in \{1, 4\}$, and $Y_i = -1$ if $X_i \in \{2, 5\}$.

(i) (6pt) Provide the pmf of $Y_1 + Y_2$ and $Y_1 Y_2$.

(ii) (6pt) For general n, find the expectation and variance of $Y_1 + \ldots + Y_n$, and an approximation expression for its cumulative distribution function. Justify your approximation.

(iii) For $c > 0$, show $P(|Y_1 \ldots Y_n| \geq c) \leq C$, where $C \downarrow 0$ as $n \to \infty$ (Hint: use Chebychev's inequality).

i) Define Y_i by $\left| \begin{array}{l} Y_i = 0 \text{ if } X_i \in \{3,6\} \\ Y_i = 1 \text{ if } X_i \in \{1,4\} \\ Y_i = -1 \text{ if } X_i \in \{2,5\} \end{array} \right]$ Each of these pairs has prob. 1/3 so $P(Y_i = 1) = P(Y_i = 0) = P(Y_i = -1) = \frac{1}{3}$, independent across i

For $S = Y_1 + Y_2$, we enumerate all pairs (Y_1, Y_2) each w/ prob. 1/9:

(Y_1, Y_2)	Sum	Prob.
$(1,1)$	2	1/9
$(1,0),(0,1)$	1	2/9
$(1,-1),(0,0),(-1,1)$	0	3/9
$(-1,0),(0,-1)$	-1	2/9
$(-1,-1)$	-2	1/9

$$\Rightarrow P(S=s) = \begin{cases} 1/9, & s = -2, 2 \\ 2/9, & s = -1, 1 \\ 1/3, & s = 0 \\ 0, & \text{otherwise} \end{cases}$$

For $P = Y_1 Y_2$, possible values are $P \in \{-1, 0, 1\}$.

$P = 0$ whenever at least one $Y_i = 0$:

$$P(P=0) = 1 - P(Y_1 \neq 0, Y_2 \neq 0) = 1 - \left(\tfrac{2}{3}\right)^2 = \tfrac{5}{9}$$

$P = 1$ whenever both are 1 or -1:

$$P(P=1) = 2 \cdot \tfrac{1}{3} \cdot \tfrac{1}{3} = \tfrac{2}{9}$$

$P = -1$ whenever one is 1 and the other is -1:

$$P(P=-1) = 2 \cdot \tfrac{1}{3} \cdot \tfrac{1}{3} = \tfrac{2}{9}$$

$$P(Y_1 Y_2 = p) = \begin{cases} 5/9, & p = 0 \\ 2/9, & p = 1 \\ 2/9, & p = -1 \end{cases}$$

ii) Let $S_n = \sum_{i=1}^{n} Y_i$; we already know $P(Y_i = 1) = P(Y_i = 0) = P(Y_i = -1) = 1/3$.

$E(Y_i) = 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + (-1) \cdot \frac{1}{3} = 0$

$\mathrm{var}(Y_i) = E(Y_i^2) - E^2(Y_i)$

where $E(Y_i^2) = 1^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + (-1)^2 \cdot \frac{1}{3} = \frac{2}{3}$

$= \frac{2}{3} - 0 = \frac{2}{3}$.

Assuming $Y_1, \ldots, Y_n \overset{iid}{\sim}$, then

$$E(S_n) = \sum_{i=1}^{n} E(Y_i) = n \cdot 0 = 0$$

$$\mathrm{var}(S_n) = \sum_{i=1}^{n} \mathrm{var}(Y_i) = n \cdot \tfrac{2}{3} = \tfrac{2n}{3}$$

Apply the CLT since $\left| \begin{array}{l} Y_i \overset{iid}{\sim} \text{ for } i = 1, \ldots, n \\ S_n = \sum_i Y_i \\ ES_n = 0,\ \mathrm{var}(S_n) = \frac{2}{3} < \infty \end{array} \right.$

Then,

$$\frac{S_n - E(S_n)}{\sqrt{\mathrm{var}(S_n)}} = \frac{S_n}{\sqrt{2n/3}} \xrightarrow{d} N(0,1).$$

Therefore, the approximate CDF is

$$P(S_n \leq t) \approx \Phi\left(\frac{t}{\sqrt{2n/3}}\right) \text{ where } \Phi := \text{standard normal cdf}$$

iii) Notice $\left| \begin{array}{l} \text{if any } Y_i = 0 \Rightarrow |Y_1 Y_2 \cdots Y_n| = 0 \\ \text{if all } Y_i \neq 0 \Rightarrow |Y_1 Y_2 \cdots Y_n| = 1 \\ \Rightarrow |Y_1 \cdots Y_n| \in \{0, 1\} \end{array} \right.$

Since $c > 0$,

$|Y_1 \cdots Y_n| \geq c \Leftrightarrow |Y_1 \cdots Y_n| = 1$ whenever $Y_i \neq 0\ \forall i$

Define $I_i = \mathbb{1}\{Y_i = 0\}$ s.t. each I_i is Bernoulli:

$P(I_i = 1) = P(Y_i = 0) = 1/3$, $P(I_i = 0) = 2/3$

$\Rightarrow E(I_i) = 1/3$, $\mathrm{var}(I_i) = \frac{1}{3}\left(\frac{2}{3}\right) = \frac{2}{9}$

Let

$Z_n = \sum_{i=1}^{n} I_i$ = number of zeros among $Y_1, \ldots, Y_n$

$\Rightarrow E(Z_n) = \frac{n}{3}$, $\mathrm{var}(Z_n) = \frac{2n}{9}$

s.t.

$|Y_1 \cdots Y_n| \geq c \Rightarrow Z_n = 0$, i.e.

$P(|Y_1 \cdots Y_n| \geq c) \leq P(Z_n = 0)$

We want to bound $P(Z_n = 0)$; observe that

$Z_n = 0 \Rightarrow |Z_n - E(Z_n)| \geq E(Z_n) = \frac{n}{3}$

so,

$$P(Z_n = 0) \leq P\left(|Z_n - E(Z_n)| \geq \tfrac{n}{3}\right).$$

Applying Chebyshev:

$$P\left(|Z_n - E(Z_n)| \geq \tfrac{n}{3}\right) \leq \frac{\mathrm{var}(Z_n)}{(n/3)^2} = \frac{2n/9}{n^2/9} = \frac{2}{n}.$$

Thus,

$$P(|Y_1 \cdots Y_n| \geq c) \leq \tfrac{2}{n}.$$

Letting $C_n = \frac{2}{n}$, then $C_n \downarrow 0$ as $n \to \infty$.

(ch. 4/5)

2. (10pt) Let U and V be i.i.d. random variables and distributed according to the uniform distribution on $[0, 1]$.

(i) (4pt) Derive an expression for the pdf of random variables $X := U + V$ and $Y := U^2$.

(ii) (3pt) Prove that X and Y are *not* independent random variables.

(iii) (3pt) Determine the joint pdf for the bivariate random vector (X, Y).

Let $\left. \begin{array}{l} U \sim \mathrm{unif}(0,1) \\ V \sim \mathrm{unif}(0,1) \end{array} \right]$ where U, V are independent

$$f_U(u) = \begin{cases} 1, & 0 < u < 1 \\ 0, & \text{otherwise} \end{cases}, \quad f_V(v) = \begin{cases} 1, & 0 < v < 1 \\ 0, & \text{otherwise} \end{cases}$$

i) $\mathrm{supp}\, f_U(u): u \in [0,1]$, $\mathrm{supp}\, f_V(v): v \in [0,1]$, $\mathrm{supp}\, X = U + V: x \in [0,2]$

Since U and V are independent with density 1 on $[0,1]$, the convolution formula gives

$$f_X(x) = \int_{\mathbb{R}} f_U(u) f_V(x-u)\, du = \int_{\mathbb{R}} f_U(u) \mathbb{1}_{\{0<u<1\}} f_V(x-u) \mathbb{1}_{\{0<x-u<1\}}$$

where $0 < x - u < 1$: $\left| \begin{array}{l} 0 < x - u \Leftrightarrow u < x \\ x - u < 1 \Leftrightarrow u > x - 1 \\ \Rightarrow x - 1 < u < x \end{array} \right.$

combining constraint from $f_U(u)$: $0 < u < 1$, then the integration interval is in the intersection:

$$\max(0, x-1) < u < \min(1, x) \text{ s.t. } f_X(x) = \int_{\max(0,x-1)}^{\min(1,x)} du$$

For lower bound: $\left| \begin{array}{l} \text{If } x - 1 < 0 \text{ (i.e. } x < 1\text{) then } \max(0, x-1) = 0 \\ \text{If } x - 1 \geq 0 \text{ (i.e. } x \geq 1\text{) then } \max(0, x-1) = x - 1 \\ \text{as lower bound changes @ } x = 1 \end{array} \right.$

upper bound: $\left| \begin{array}{l} \text{If } x < 1, \text{ then } \min(1, x) = x \\ \text{If } x > 1, \text{ then } \min(1, x) = 1 \\ \text{where upper bound also changes @ } x = 1 \end{array} \right.$

Since both bounds switch form @ $x = 1$, split integral into cases:

$$0 < x < 1: \ f_X(x) = \int_{\max(0,x-1)}^{\min(1,x)} du = \int_0^x du = x$$

$$1 \leq x < 2: \ f_X(x) = \int_{x-1}^{1} du = u \Big|_{x-1}^{1} = 1 - (x-1) = 2 - x$$

$$\Rightarrow f_X(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 \leq x < 2 \\ 0, & \text{otherwise} \end{cases}$$

Now, for $Y = U^2$, $\mathcal{Y} = [0^2, 1^2] = [0, 1]$,

$$F_Y(y) = P(U^2 \leq y) = P(U \leq \sqrt{y}) \text{ where } U \sim \mathrm{unif}(0,1)$$
$$= \sqrt{y}$$

$$f_Y(y) = F_Y'(y) = \frac{d}{dy} y^{\frac{1}{2}} = \frac{1}{2} y^{-\frac{1}{2}} = \frac{1}{2\sqrt{y}}, \ 0 < y < 1$$

$$\text{Thus, } f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}, & 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

ii) We know that $\left| \begin{array}{l} X = U + V \\ Y = U^2 \\ U, V \sim \mathrm{Unif}(0,1) \text{ are ind.} \end{array} \right.$

If $Y = y$, then necessarily $U = \sqrt{y}$ (since $U \geq 0$); thus,

$$X = \sqrt{y} + V \in [\sqrt{y}, \sqrt{y} + 1]$$

Thus, the joint support of (X, Y) is contained in:

$$\{(x, y): 0 < y < 1, \sqrt{y} < x < \sqrt{y} + 1\}.$$

Choose $y = 1/4$ and $x = 0.1$. Then

$$x < \sqrt{y} = 0.5 \Rightarrow P(X = x, Y = y) = 0.$$

However, $f_X(0.1) > 0$ and $f_Y(1/4) > 0$ since both X and Y have positive densities on their supports.

Therefore

$$f_{XY}(x, y) \neq f_X(x) f_Y(y) \Rightarrow X \text{ and } Y \text{ not independent.}$$

iii) Determine the joint pdf of (X, Y) where

transformation: $\left| \begin{array}{l} X = U + V \\ Y = U^2 \end{array} \right.$ and inverse transformation: $\left| \begin{array}{l} U = \sqrt{Y} \\ V = X - \sqrt{Y} \end{array} \right.$

Then

$$J = \frac{\partial(u, v)}{\partial(x, y)} = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} = \begin{pmatrix} 0 & \frac{1}{2\sqrt{y}} \\ 1 & -\frac{1}{2\sqrt{y}} \end{pmatrix} = \frac{1}{2\sqrt{y}}$$

So,

$$f_{XY}(x, y) = f_{U,V}(\sqrt{y}, x - \sqrt{y}) \cdot |\det(J)| = \frac{1}{2\sqrt{y}}$$

where $0 < u < 1$ and $0 < v < 1 \Rightarrow \left| \begin{array}{l} 0 < \sqrt{y} < 1 \Rightarrow 0 < y < 1 \\ 0 < x - \sqrt{y} < 1 \Rightarrow \sqrt{y} < x < \sqrt{y} + 1 \end{array} \right.$

and $f_{UV}(u, v) = f_U(u) f_V(v)$ via independence where $U, V \sim \mathrm{unif}(0,1)$

$= 1$ for $0 < u < 1$, $0 < v < 1$

$$f_{XY}(x, y) = \begin{cases} \frac{1}{2\sqrt{y}}, & 0 < y < 1, \sqrt{y} < x < \sqrt{y} + 1 \\ 0, & \text{otherwise} \end{cases}$$

(ch. 3)

3. (6pt) Which of the following classes of distributions belong to the exponential families (justify!)

(i) discrete distributions on integers;
(ii) the negative binomial distributions, $P(X=x|r,p) = \binom{x-1}{r-1}p^r(1-p)^{x-r}$, $x = r, r+1, \ldots$;
(iii) mixture distributions of two or more Gamma distributions, i.e., those with the pdf of the form $f(x|k,p,\alpha,\beta) = \sum_{i=1}^{k} p_i \mathrm{Gamma}(x|\alpha_i,\beta_i)$, for some $k \in \mathbb{N}$, $\alpha_i, \beta_i > 0$ for $i = 1, \ldots, k$.

Recall: A family of pmf or pdf is called an exponential family if it has the form:

$$f(x|\theta) = c(\theta)h(x)\exp\left\{\sum_{i=1}^{k} w_i(\theta)t_i(x)\right\}, \quad x \in \mathcal{X}$$
$$= c(\theta)h(x)\exp\langle w(\theta), t(x)\rangle$$

where $h(x) \geq 0$
$t(x) = (t_1(x), \ldots, t_k(x))$ depends only on x
$w(\theta) = (w_1(\theta), \ldots, w_k(\theta))$ depends only on θ

i) Discrete distributions on the integers

The class of all discrete distributions on $\mathbb{Z}$ (integers) consists of arbitrary pmf's:

$$P(X=k) = p_k, \quad k \in \mathbb{Z}, \quad \sum_k p_k = 1$$

This class requires infinitely many free parameters $\{p_k\}$.

By definition, an exponential family must be parameterized by a finite-dimensional parameter vector θ. Therefore, the class of all discrete distributions on the integers cannot be represented in exponential family form.

ii) Given $P(X=x|r,p) = \binom{x-1}{r-1}p^r(1-p)^{x-r}$, $x = r, r+1, \ldots$, r: fixed and $0<p<1$

$$= \binom{x-1}{r-1}p^r(1-p)^{-r}(1-p)^x$$

where $a^x = \exp(x\log a)$

$$= \underbrace{p^r(1-p)^{-r}}_{c(p)}\underbrace{\binom{x-1}{r-1}}_{h(x)}\exp(\underbrace{\log(1-p)}_{w(p)}\cdot\underbrace{x}_{t(x)}).$$

whereas r: unknown means $(x-r)!$ term is inseparable w.r.t. parameter: r and data: x

iii) Mixture distributions of two/more Gamma distributions, i.e.

$$f(x|k,p,\alpha,\beta) = \sum_{i=1}^{k} p_i \,\mathrm{Gamma}(x, \alpha_i, \beta_i) \text{ for some } k \in \mathbb{N}, \alpha_i, \beta_i > 0 \text{ for } i = 1, \ldots, k$$

Each Gamma distribution individually belongs to the exponential family:

$$f_i(x) = h(x)\exp\left(\sum_{j=1}^{m} w_{ij}(\theta)t_j(x)\right)$$

whereas a mixture is

$$f(x) = \sum_{i=1}^{k} p_i f_i(x) \quad \text{assuming } \sum_{i=1}^{k} p_i = 1$$
$$= h(x)\sum_{i=1}^{k} p_i \underbrace{\exp\left(\sum_{j=1}^{m} w_{ij}(\theta)t_j(x)\right)}$$

which violates the single-exponential structure since there is no algebraic identity that turns $\sum_i p_i e^{a_i(x)}$ into $e^{a(x)}$

In general, mixtures introduce sums of exponentials, not exponentials of sums.

(ch. 4(5))

4. (8pt) Suppose that a signal of interest is represented by a random variable X and we assume that $X \sim \mathrm{Normal}(\mu, \sigma^2)$. The signal is contaminated by an amount of noise ϵ, so what is observed is $Y = X + \epsilon$. The noise ϵ is not independent of X; in fact, $\epsilon = aX + Z$, where a is a constant, Z is independent of X and $Z \sim \mathrm{Normal}(0, 1)$.

(i) (4pt) Argue that (X, Y) is a bivariate normal vector. What is its mean and covariance matrix?
(ii) (2pt) Suppose that $corr(X, Y) < 0$. What does that say about the constant a?
(ii) (2pt) Given an observation $Y = y$, derive $\mathbb{E}[X|Y = y]$.

Let $X \sim N(\mu, \sigma^2)$

$\epsilon = aX + Z$, a: constant, $Z \sim N(0,1)$ independent of X

The observed signal is $Y = X + \epsilon$

i) Argue that (X,Y) is a bivariate normal vector. What is its mean and covariance matrix.

Rewriting Y:

$$Y = X + \epsilon = X + aX + Z = X(1+a) + Z$$
$$\Leftrightarrow$$
$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1+a & 1 \end{pmatrix}\begin{pmatrix} X \\ Z \end{pmatrix}.$$

Since $X \sim N(\mu, \sigma^2)$, $Z \sim N(0,1)$, X and Z are independent, $(X,Z)^T$ is jointly normal and any linear transformation of a jointly normal vector is also jointly normal.
Therefore
$(X,Y)^T$ is bivariate normal.

$$E\left[\begin{pmatrix} X \\ Y \end{pmatrix}\right] = \begin{bmatrix} E(X) \\ E(Y) \end{bmatrix} \text{ where } \begin{array}{|l} EX = \mu \\ EY = E[(1+a)X + Z] = (1+a)\mu \end{array}$$

compute the covariance matrix: $\mathrm{var}X = \sigma^2$
$\mathrm{var}Z = 1$
$\mathrm{cov}(X,Z) = E[(X-EX)(Z-EZ)]$
$= E(XZ) - E(X)E(Z)$
$= E(X)E(Z) - EXEZ$ via independence
$= 0$

Next,
$\mathrm{var}Y = \mathrm{var}((1+a)X + Z) = (1+a)^2\sigma^2 + 1$
$\mathrm{cov}(X,Y) = \mathrm{cov}(X, (1+a)X + Z)$
★ via linearity of covariance: $\mathrm{cov}(X, A+B) = \mathrm{cov}(X,A) + \mathrm{cov}(X,B)$
$= \mathrm{cov}(X, (1+a)X) + \mathrm{cov}(X,Z)$
★ $\mathrm{cov}(X, cX) = c\,\mathrm{var}X$ ∀ constant c
$= (1+a)\mathrm{var}X.$

Therefore,

$$\Sigma = \begin{pmatrix} \mathrm{var}X & \mathrm{cov}(X,Y) \\ \mathrm{cov}(X,Y) & \mathrm{var}Y \end{pmatrix} = \begin{pmatrix} \sigma^2 & (1+a)\sigma^2 \\ (1+a)\sigma^2 & (1+a)^2\sigma^2 + 1 \end{pmatrix}.$$

ii) Suppose $\mathrm{corr}(X,Y) < 0$.
What does this say about the constant a.

Recall that
$$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}}$$

and from (i):
$$\mathrm{cov}(X,Y) = (1+a)\sigma^2.$$

Since $\sigma^2 > 0$ and the denominator is positive, the sign of the correlation is determined by $(a+1)$.

Thus,
$$\mathrm{corr}(X,Y) < 0 \Leftrightarrow (1+a) < 0 \Leftrightarrow \underline{a < -1}.$$

iii) Given an observation $Y=y$, derive $E[X|Y=y]$

Fact: If (X,Y) is jointly normal, then
$E[X|Y=y]$ is a linear function of y
$\Rightarrow E[X|Y=y] = \alpha + \beta y$ for some constants α, β.

Find β via

★ property: $\mathrm{cov}(X,Y) = \mathrm{cov}(E(X|Y), Y)$

$\mathrm{cov}(X,Y) = \mathrm{cov}(\alpha + \beta Y, Y)$
where $\mathrm{cov}(\alpha, Y) = 0$ and $\mathrm{cov}(\beta Y, Y) = \beta\,\mathrm{var}Y$
$= \beta\,\mathrm{var}Y$

$$\Rightarrow \beta = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}Y}$$

Find α using expectations:

$$E[E[X|Y]] = E(X) = E(\alpha + \beta Y) = \alpha + \beta E(Y)$$

$$\Rightarrow \alpha = EX - \beta EY$$

Substituting α, β back:

$$E[X|Y=y] = E(X) - \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(Y)}E(Y) + \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(Y)}Y$$
$$= E(X) + \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(Y)}(Y - EY)$$
$$= \mu + \frac{(1+a)\sigma^2}{(1+a)^2\sigma^2 + 1}(Y - (1+a)\mu)$$

(ch. 4)

5. (10pt) Let p, Z, X be random variables whose joint distribution is given by the hierarchical model:

$$p \sim \mathrm{Beta}(\alpha, \beta), \quad (1)$$
$$Z|p \sim \mathrm{Bernoulli}(p) \quad (2)$$
$$X|p, Z=0 \sim \mathrm{Normal}(\mu_0, 1) \quad (3)$$
$$X|p, Z=1 \sim \mathrm{Normal}(\mu_1, 1), \quad (4)$$

where parameters $\alpha, \beta > 0$, $\mu_0, \mu_1 \in \mathbb{R}$.

(i) (5pt) Derive the pmf/pdf for the (marginal) distribution for the variable Z, and the conditional distribution of X given Z (that is, the conditional pdf of X given $Z = z$ for $z \in \{0, 1\}$).
(ii) (3pt) Find the (marginal) expectation of X. (Hint: use iterated expectation formula).
(iii) (2pt) Derive the conditional distribution for Z given X (that is, find the expression for $P(Z = z|X = x)$ for $z \in \{0, 1\}$).

Let p, Z, X be RV's whose joint distribution is given by the hierarchical model:
$p \sim \mathrm{Beta}(\alpha, \beta)$ (1)
$Z|p \sim \mathrm{Bernoulli}(p)$ (2)
$X|p, Z=0 \sim \mathrm{Normal}(\mu_0, 1)$ (3)
$X|p, Z=1 \sim \mathrm{Normal}(\mu_1, 1)$ (4)

where parameters $\alpha, \beta > 0$, $\mu_0, \mu_1 \in \mathbb{R}$.

i) Derive the pmf/pdf for the (marginal) distribution for the variable Z and the conditional distribution of $X|Z$ (i.e. conditional pdf of $X|Z=z$ for $z \in \{0,1\}$)

ia) marginal pmf of Z:

starting from the hierarchy: $p \sim \mathrm{Beta}(\alpha, \beta)$, $Z|p \sim \mathrm{Bernoulli}(p)$

we want $P(Z=z)$ for $z \in \{0,1\}$.

using the following identity: ∀ event A,
★ $P(A) = E(1_A)$ where $1_A = \begin{cases} 1, & A \text{ occurs} \\ 0, & \text{otherwise} \end{cases}$

For $z=1$: $P(Z=1) = E[P(Z=1|p)] = E(p) = \frac{\alpha}{\alpha+\beta}$
$P(Z=0) = 1 - P(Z=1) = \frac{\alpha+\beta}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta} = \frac{\beta}{\alpha+\beta}$

and therefore,

$$P(Z=z) = \begin{cases} \frac{\beta}{\alpha+\beta}, & z=0 \\ \frac{\alpha}{\alpha+\beta}, & z=1 \end{cases} \Rightarrow Z \sim \mathrm{Bernoulli}\left(\frac{\alpha}{\alpha+\beta}\right).$$

ib) conditional pdf of $X|Z=z$:

$X|p, Z=0 \sim N(\mu_0, 1)$, $X|p, Z=1 \sim N(\mu_1, 1)$

Notice that, once Z is given, the distribution of x does not depend on p.
So, integrating out p does nothing here.

If $X \sim N(\mu, \sigma^2)$, then
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right)$$

writing out the conditional pdf's:

for $z=0$: $f_{X|Z=0}(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-\mu_0)^2}{2}\right)$

for $z=1$: $f_{X|Z=1}(x) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-(x-\mu_1)^2}{2}\right)$

ii) Find the (marginal) expectation of X (Hint: use iterated expectation formula), i.e.
obtain $E(X)$ using the identity: $E(X) = E[E(X|Z)]$.

Given $X|Z=0 \sim N(\mu_0, 1)$, $X|Z=1 \sim N(\mu_1, 1)$ $\Rightarrow$ $E(X|Z=0) = \mu_0$, $E(X|Z=1) = \mu_1$ $\Rightarrow E(X|Z) = \begin{cases} \mu_0, & Z=0 \\ \mu_1, & Z=1 \end{cases}$

Now, since Z is discrete, applying our outer expectation,

$$E(X) = \mu_0 P(Z=0) + \mu_1 P(Z=1)$$
$$= \mu_0\frac{\beta}{\alpha+\beta} + \mu_1\frac{\alpha}{\alpha+\beta}$$
$$\Rightarrow E(X) = \frac{\beta\mu_0 + \alpha\mu_1}{\alpha+\beta}$$

iii) Derive the conditional distribution for Z given X, i.e.
find expression for $P(Z=z|X=x)$ for $z \in \{0,1\}$.
Z is discrete with values $\{0,1\}$.
So,
$Z|X=x$ is described by a pmf: $P(Z=0|X=x)$, $P(Z=1|X=x)$
Using Bayes' Rule,

$$\forall z \in \{0,1\}, \quad P(Z=z|X=x) = \frac{f_{X|Z}(x|z)P(Z=z)}{f_X(x)}$$

From (i): $P(Z=1) = \frac{\alpha}{\alpha+\beta}$, $P(Z=0) = \frac{\beta}{\alpha+\beta}$

Letting $\phi(t) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)$,

$$f_{X|Z=z}(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-\mu_z)^2}{2}\right) =: \phi(x-\mu_z).$$

After substitution,

$$P(Z=1|X=x) = \frac{f_{X|Z=1}(x)P(Z=1)}{f_X(x)} = \frac{\frac{\alpha}{\alpha+\beta}\phi(x-\mu_1)}{\frac{\beta}{\alpha+\beta}\phi(x-\mu_0) + \frac{\alpha}{\alpha+\beta}\phi(x-\mu_1)} = \frac{\alpha\phi(x-\mu_1)}{\beta\phi(x-\mu_0) + \alpha\phi(x-\mu_1)}$$

Similarly, $P(Z=0|X=x) = 1 - P(Z=1|X=x) = \dfrac{\beta\phi(x-\mu_1)}{\beta\phi(x-\mu_0) + \alpha\phi(x-\mu_1)}$.

Therefore,

$$\underline{Z|X=x \sim \mathrm{Bernoulli}\left(\frac{\alpha\phi(x-\mu_1)}{\beta\phi(x-\mu_0) + \alpha\phi(x-\mu_1)}\right)}$$