HIERARCHICAL MODELS

When considering two r.v.s X and Y, we may in some situations possess insight about the conditional distribution of $Y|X$ and the marginal distribution of X, while we may, in the end, be primarily interested in the marginal distribution of Y. Such is an occasion for **hierarchical models**, of which we will discuss some examples.

Poisson-Binomial hierarchical model:

Let X = # customers entering a store in a day
Y = # customers who will make purchases in the store in a day.

Suppose it is Y in which we are primarily interested.

We might assume the following hierarchical model for Y:

$$Y|X \sim \text{Binomial}(X, p) \quad \text{and} \quad X \sim \text{Poisson}(\lambda).$$

This model assumes that the number Y of customers who make purchases depends upon the number X of customers to enter the store, such that each of the X customers to enter makes a purchase with probability p and the customers act independently. Moreover, the number of customers to enter the store is assumed to have the Poisson (λ) distribution.

The marginal pmf of Y is obtained by

(i) finding the joint pmf of X and Y

(ii) summing this over all values of X.

We have the conditional and marginal pmfs

$$p(y|x) = \binom{x}{y} p^y (1-p)^{x-y}, \qquad y = 0, 1, \ldots, x$$

$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots,$$

giving the joint pmf

The joint pmf is the product of the conditional and marginal pmfs

$$p(x,y) = \binom{x}{y} p^y (1-p)^{x-y} \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots, \quad y = 0, 1, \ldots, x.$$

So the marginal pmf of Y is

$$p_Y(y) = \sum_{x=y}^{\infty} \binom{x}{y} p^y (1-p)^{x-y} \frac{e^{-\lambda}\lambda^x}{x!}, \quad y = 0,1,2,\ldots$$

$x \geq y$

$$= \sum_{x=y}^{\infty} \frac{(p\lambda)^y ((1-p)\lambda)^{x-y}}{y!(x-y)!} e^{-\lambda}, \quad y = 0,1,2,\ldots$$

$$= \frac{(p\lambda)^y}{y!} e^{-p\lambda} \sum_{x=y}^{\infty} \frac{((1-p)\lambda)^{x-y} e^{-(1-p)\lambda}}{(x-y)!}, \quad y = 0,1,2,\ldots$$

$$= \frac{(p\lambda)^y e^{-p\lambda}}{y!}, \quad y = 0,1,2,\ldots$$

So the marginal distribution of Y is the Poisson$(p\lambda)$ distribution.

Therefore we have $\mathbb{E}Y = p\lambda$ and $\mathrm{Var}\, Y = p\lambda$.

The following results can make it easier to compute unconditional expected values and variances, particularly in the context of hierarchical models:

Thm: For any r.v.s X and Y

(i) $\mathbb{E}Y = \mathbb{E}\left(\mathbb{E}[Y|X]\right)$ (This is sometimes called "iterated" expectation)

(ii) $\mathrm{Var}\, Y = \mathbb{E}\left(\mathrm{Var}[Y|X]\right) + \mathrm{Var}\left(\mathbb{E}[Y|X]\right)$,

provided that the expectations exist.

Proof: (i) Suppose X and Y are continuous r.v.s.

$$\mathbb{E}Y = \int_{\mathbb{R}} y\, f_Y(y)\, dy = \int_{\mathbb{R}} y \left[\int_{\mathbb{R}} f(x,y)\, dx\right] dy$$

$$f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}$$

$$\iff f_{Y|X}(y|x) f_X(x) = f_{(X,Y)}(x,y)$$

$$= \int_{\mathbb{R}} y \left[\int_{\mathbb{R}} f(y|x) f_X(x)\, dx\right] dy$$

$$= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} y\, f(y|x)\, dy\right] f_X(x)\, dx$$

$$= \int_{\mathbb{R}} \mathbb{E}[Y|X=x] f_X(x)\, dx$$

$$= \mathbb{E}\left(\mathbb{E}[Y|X]\right).$$

For (X,Y) discrete, replace pdfs with pmfs and integrals with sums.

(ii) $$\operatorname{Var} Y = \mathbb{E}\left(Y - \mathbb{E}Y\right)^2$$

$$= \mathbb{E}\left(Y - \mathbb{E}\left(\mathbb{E}[Y|X]\right)\right)^2$$

$$= \mathbb{E}\Big(Y \underbrace{- \mathbb{E}[Y|X] + \mathbb{E}[Y|X]}_{\text{add and subtract } \mathbb{E}[Y|X]} - \mathbb{E}\left(\mathbb{E}[Y|X]\right)\Big)^2$$

$$= \mathbb{E}\left(Y - \mathbb{E}[Y|X]\right)^2 + \mathbb{E}\left(\mathbb{E}[Y|X] - \mathbb{E}\left(\mathbb{E}[Y|X]\right)\right)^2$$

iterate the expectation!

$$+ 2\,\mathbb{E}\left(Y - \mathbb{E}[Y|X]\right)\underbrace{\mathbb{E}\left(\mathbb{E}[Y|X] - \mathbb{E}\left(\mathbb{E}[Y|X]\right)\right)}_{\mathbb{E}(\mathbb{E}[Y|X]) - \mathbb{E}(\mathbb{E}[Y|X]) = 0}$$

$$= \mathbb{E}\Big(\underbrace{\mathbb{E}\left[\left(Y - \mathbb{E}[Y|X]\right)^2 \middle| X\right]}_{\operatorname{Var}[Y|X]}\Big) + \mathbb{E}\left(\mathbb{E}[Y|X] - \mathbb{E}\left(\mathbb{E}[Y|X]\right)\right)^2$$

$$= \mathbb{E}\left(\operatorname{Var}[Y|X]\right) + \operatorname{Var}\left(\mathbb{E}[Y|X]\right)$$

Poisson-Binomial hierarchical model revisited:

Let $Y|X \sim \text{Binomial}(X, p)$ and $X \sim \text{Poisson}(\lambda)$.

Then $\mathbb{E}[Y|X] = Xp$ and $\operatorname{Var}[Y|X] = Xp(1-p)$

and $\mathbb{E}X = \lambda$ and $\operatorname{Var} X = \lambda$.

So we get $\mathbb{E}Y = \mathbb{E}\left(\mathbb{E}[Y|X]\right) = \mathbb{E}(Xp) = p\,\mathbb{E}X = p\lambda$

and
$$\operatorname{Var} Y = \mathbb{E}\left(\operatorname{Var}[Y|X]\right) + \operatorname{Var}\left(\mathbb{E}[Y|X]\right)$$

$$= \mathbb{E}\left(Xp(1-p)\right) + \operatorname{Var}(Xp)$$

$$= p(1-p)\,\mathbb{E}X + p^2 \operatorname{Var} X$$

$$= p(1-p)\lambda + p^2\lambda$$

$$= p\lambda.$$

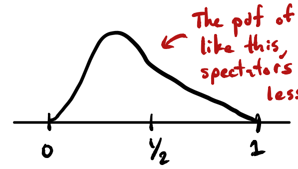
Beta-Binomial hierarchical model:

Suppose Y = # free-throws made out of n attempts of a randomly selected spectator at a basketball game. Let P = the free-throw success rate of a randomly selected spectator.

We might assume the following hierarchical model for Y:

$$Y|P \sim \text{Binomial}(n, P) \quad \text{and} \quad P \sim \text{Beta}(\alpha, \beta).$$

We can find the mean and variance of Y as follows:



The pdf of P could look like this, under which most spectators have success rate less than $\frac{1}{2}$.

$$\mathbb{E}Y = \mathbb{E}\left(\mathbb{E}[Y|P]\right) = \mathbb{E}(nP) = n\,\mathbb{E}P = n\left(\frac{\alpha}{\alpha+\beta}\right)$$

$$\begin{aligned}
\text{Var}\,Y &= \mathbb{E}\left(\text{Var}[Y|P]\right) + \text{Var}\left(\mathbb{E}[Y|P]\right)\\
&= \mathbb{E}\left(nP(1-P)\right) + \text{Var}(nP)\\
&= n\,\mathbb{E}P - n\,\mathbb{E}P^2 + n^2\,\text{Var}\,P\\
&= n\,\mathbb{E}P - n\left(\text{Var}\,P + (\mathbb{E}P)^2\right) + n^2\,\text{Var}\,P\\
&= n\left(\frac{\alpha}{\alpha+\beta}\right) - n\left(\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} + \frac{\alpha^2}{(\alpha+\beta)^2}\right) + n^2\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\\
&= n\left[\frac{\alpha(\alpha+\beta)(\alpha+\beta+1) - \alpha\beta - \alpha^2(\alpha+\beta+1)}{(\alpha+\beta)^2(\alpha+\beta+1)}\right] + n^2\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\\
&= n\left[\frac{\alpha\beta(\alpha+\beta+1) - \alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\right] + n^2\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\\
&= n\,\frac{\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}.
\end{aligned}$$

The marginal pmf of Y is given by

$$p_Y(y) = \int_0^1 \underbrace{\binom{n}{y}p^y(1-p)^{n-y}}_{p_{Y|P}(y|p)}\;\underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}}_{f_P(p)}\,dp$$

$$= \int_0^1 \frac{n!}{(n-y)!\,y!} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha+y-1}(1-p)^{n-y+\beta-1}\,dp$$

$$= \frac{n!}{(n-y)!\,y!} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+y)\,\Gamma(n-y+\beta)}{\Gamma((\alpha+y)+(n-y+\beta))}$$

$$\times \underbrace{\int_0^1 \frac{\Gamma((\alpha+y)+(n-y+\beta))}{\Gamma(\alpha+y)\,\Gamma(n-y+\beta)} p^{\alpha+y-1}(1-p)^{n-y+\beta-1}\,dp}_{=1}$$

$$= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+y)\,\Gamma(n-y+\beta)}{\Gamma(\alpha+\beta+n)}, \quad \text{for } y=0,1,\dots,n$$

This is indeed a pmf (sums to 1 over $y=0,1,\dots,n$); it is the pmf of a distribution called the Beta-binomial distribution.

Normal "random effects" model:

Suppose Y is the score on a standardized test of a randomly selected pupil in the U.S. Suppose A is the average test score at the school of the selected pupil.

Then we might assume the following hierarchical model for Y:

$Y|A \sim \text{Normal}(A, \sigma^2)$ and $A \sim \text{Normal}(\mu_A, \sigma_A^2)$

Assume that pupils' scores are Normally distributed about their school average with variance σ^2.

Assume that school averages are Normal with mean μ_A and variance σ_A^2.

Then the expectation and variance of Y are

$$\mathbb{E}Y = \mathbb{E}(\mathbb{E}[Y|A]) = \mathbb{E}(A) = \mu_A$$

$$\text{Var}\,Y = \mathbb{E}(\text{Var}[Y|A]) + \text{Var}(\mathbb{E}[Y|A]) = \mathbb{E}(\sigma^2) + \text{Var}(A) = \underbrace{\sigma^2 + \sigma_A^2}.$$

This hierarchical model for Y allows us to decompose the variance of Y into that among pupils at a given school, σ^2, and that between schools, σ_A^2.

The marginal pdf of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left[-\frac{1}{2}\frac{(y-a)^2}{\sigma^2}\right]}_{f_{Y|A}(y|a)} \underbrace{\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_A}\exp\left[-\frac{1}{2}\frac{(a-\mu_A)}{\sigma_A^2}\right]^2}_{f_A(a)} da$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma\sigma_A} \exp\left[-\frac{1}{2}\frac{\sigma_A^2(y-a)^2 + \sigma^2(a-\mu_A)^2}{\sigma^2\sigma_A^2}\right] da$$

See Addendum for these algebraic details

$$= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\sigma^2+\sigma_A^2}}\exp\left[-\frac{1}{2}\frac{(y-\mu_A)^2}{\sigma_A^2+\sigma^2}\right] \underbrace{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\left(\frac{\sigma\sigma_A}{\sqrt{\sigma^2+\sigma_A^2}}\right)^{-1}\exp\left[-\frac{1}{2}\frac{\left(a-\left(\frac{\sigma_A^2 y-\sigma^2\mu_A}{\sigma_A^2+\sigma^2}\right)\right)^2}{\sigma^2\sigma_A^2/(\sigma_A^2+\sigma^2)}\right] da}_{=1}$$

Integral over pdf of Normal$\left(\frac{\sigma_A^2 y+\sigma^2\mu_A}{\sigma_A^2+\sigma^2}, \frac{\sigma^2\sigma_A^2}{\sigma_A^2+\sigma^2}\right)$ distribution

$$= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\sigma^2+\sigma_A^2}}\exp\left[-\frac{1}{2}\frac{(y-\mu_A)^2}{\sigma_A^2+\sigma^2}\right].$$

So $Y \sim \text{Normal}(\mu_A, \sigma_A^2+\sigma^2)$.

Note also that the above gives $A|Y \sim \text{Normal}\left(\frac{\sigma_A^2+\sigma^2\mu_A}{\sigma_A^2+\sigma^2}, \frac{\sigma^2\sigma_A^2}{\sigma_A^2+\sigma^2}\right)$.

Remark: In this example we found $f_Y(y)$ by doing these steps:

$$f_Y(y) = \int_{\mathbb{R}} f(y,a)\, da$$

Write the joint pdf as the product of the conditional and marginal pdfs

$$= \int_{\mathbb{R}} f(y|a) f_A(a)\, da$$

Rearrange to get a function of only y times a function of a and y

$$= \int_{\mathbb{R}} g(a,y)\, h(y)\, da$$

Take the function of only y outside the integral.

$$= h(y)\int_{\mathbb{R}} g(a,y)\, da$$

Adjust constants, changing g, h to g', h' so that $\int_{\mathbb{R}} h'(y)\,dy = 1$ and $\int_{\mathbb{R}} g'(a,y)\,da = 1$

$$= h'(y)\underbrace{\int_{\mathbb{R}} g'(a,y)\, da}_{=1}$$

Then $h' = f_Y(y)$ and $g' = f(a|y)$.

"MULTINOULLI" DISTRIBUTION AND MIXTURE OF GAUSSIANS

In order to present the next hierarchical model we need to introduce the "Multinoulli" trial, which is an extension of the Bernoulli trial that allows more than two outcomes.

A "Multinoulli" trial is an experiment in which there are K outcomes occurring with probabilities $p_1, \dots, p_K$, where $\sum_{k=1}^{K} p_k = 1$.

(Recall the Bernoulli trial, in which there are two outcomes called "success" and "failure" which occur with probabilities p and $1-p$.)

Let the r.v.s $X_1, \dots, X_K$ encode the outcome of a multinoulli trial as

$$X_k = \begin{cases} 1 & \text{if outcome } k \text{ occurs} \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } k = 1, \dots, K.$$

Then $(X_1, \dots, X_K)$ has the Multinoulli$(p_1, \dots, p_K)$ distribution, and the joint pmf of $(X_1, \dots, X_K)$ is given by

$$p(x_1, \dots, x_K) = p_1^{x_1} \cdot \ldots \cdot p_K^{x_K} \cdot \mathbb{1}\Big(\underbrace{(x_1, \dots, x_K) \in \{0,1\}^K}_{\text{support of } X_1, \dots, X_K}, \ \sum_{k=1}^{K} x_k = 1 \Big).$$

only one of $x_1, \dots, x_K$ may equal 1.

E.g. A customer redeems coupon for (1) ice cream with probability $1/4$,
(2) French fries with probability $2/4$,
(3) a donut with probability $1/4$.

Let (X_1, X_2, X_3) be the triplet of r.v.s

$$(X_1, X_2, X_3) = \begin{cases} (1,0,0) & \text{if ice cream} \\ (0,1,0) & \text{if French fries} \\ (0,0,1) & \text{if a donut} \end{cases}$$

Then $p(1,0,0) = 1/4$, $p(0,1,0) = 2/4$, $p(0,0,1) = 1/4$

We have for $(X_1, \dots, X_K) \sim$ Multinoulli$(p_1, \dots, p_K)$ that the marginal pmf of X_1 is

$$p_{X_1}(x_1) = \sum_{\{(x_2, \dots, x_K) \in \{0,1\}^{K-1} : \sum_{k=2}^{K} x_k = 1 - x_1\}} p_1^{x_1} \cdot p_2^{x_2} \cdot \ldots \cdot p_K^{x_K}$$

$$= \begin{cases} p_1 & \text{if } x_1 = 1 \\ p_2 + p_3 + \dots + p_K = 1 - p_1 & \text{if } x_1 = 0, \end{cases}$$

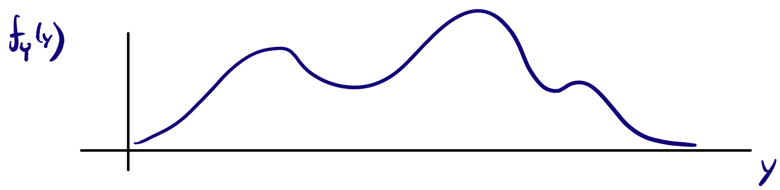
so the marginal distribution of X_k is the Bernoulli(p_k) dist. for $k = 1, \dots, K$.

Moreover, $\mathrm{Cov}(X_k, X_{k'}) = \mathbb{E}\underbrace{X_k X_{k'}}_{=0 \text{ if } k \neq k'} - \mathbb{E}X_k \mathbb{E}X_{k'} = \begin{cases} -p_k p_{k'} & \text{if } k \neq k' \\ p_k(1-p_k) & \text{if } k = k'. \end{cases}$

Mixture of Gaussians:

The Normal distribution is also called the Gaussian distribution after Carl Friedrich Gauß

If we believe a r.v. Y has a pdf with multiple modes (local maxima), we might assume that the pdf is composed of several densities added together.



We might assume the following hierarchical model for Y:

$$Y \mid (X_1, \dots, X_K) \sim \mathrm{Normal}\left(\sum_{k=1}^{K} X_k \mu_k, \sum_{k=1}^{K} X_k \sigma_k^2 \right),$$

for some $\mu_1, \dots, \mu_K, \sigma_1^2, \dots, \sigma_K^2$, with

$$(X_1, \dots, X_K) \sim \mathrm{Multinoulli}(p_1, \dots, p_K),$$

so $Y \mid (X_1, \dots, X_K)$ has the $\mathrm{Normal}(\mu_k, \sigma_k^2)$ distribution if $X_k = 1$, i.e. if outcome k occurs.

Then the expectation and variance of Y are

$$\mathbb{E}Y = \mathbb{E}\left(\mathbb{E}[Y \mid (X_1, \dots, X_K)] \right) = \mathbb{E}\left(\sum_{k=1}^{K} X_k \mu_k \right) = \underbrace{\sum_{k=1}^{K} p_k \mu_k}_{\text{A linear combination of } \mu_1, \dots, \mu_K \text{ with the weights } p_1, \dots, p_K}$$

$$\begin{aligned} \mathrm{Var}\, Y &= \mathbb{E}\left(\mathrm{Var}[Y \mid (X_1, \dots, X_K)] \right) + \mathrm{Var}\left(\mathbb{E}[Y \mid (X_1, \dots, X_K)] \right) \\ &= \mathbb{E}\left(\sum_{k=1}^{K} X_k \sigma_k^2 \right) + \mathrm{Var}\left(\sum_{k=1}^{K} X_k \mu_k \right), \end{aligned}$$

where

$$\mathbb{E}\left(\sum_{k=1}^{K} X_k \sigma_k^2 \right) = \sum_{k=1}^{K} p_k \sigma_k^2$$

and

$$\mathrm{Var}\left(\sum_{k=1}^{K} X_k \mu_k \right) = \sum_{k=1}^{K} \mu_k^2 \mathrm{Var}(X_k) + 2 \sum_{k > k'} \mu_k \mu_{k'} \mathrm{Cov}(X_k, X_{k'})$$

$$= \sum_{k=1}^{K} \mu_k^2 p_k(1-p_k) - 2\sum_{k>k'} \mu_k \mu_{k'} p_k p_{k'}$$

$$= \sum_{k=1}^{K} \mu_k^2 p_k - \sum_{k=1}^{K}\sum_{k'=1}^{K} \mu_k \mu_{k'} p_k p_{k'}$$

$$= \sum_{k=1}^{K} \mu_k^2 p_k - \left(\sum_{k=1}^{K} \mu_k p_k\right)^2$$

$$= \sum_{k=1}^{K} p_k \left(\mu_k - \left(\sum_{k=1}^{K} \mu_k p_k\right)\right)^2,$$

so that

$$\operatorname{Var} Y = \underbrace{\sum_{k=1}^{K} p_k \sigma_k^2}_{\text{weighted mean of variances}} + \underbrace{\sum_{k=1}^{K} p_k \left(\mu_k - \left(\sum_{k=1}^{K} \mu_k p_k\right)\right)^2}_{\text{weighted variance of means}}.$$

The marginal pdf of Y is given by

$$f_Y(y) = \sum_{\{(x_1,\dots,x_K)\in\{0,1\}^K : \sum_{k=1}^{K} x_k = 1\}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sum x_k \sigma_k^2}} \exp\left[-\frac{1}{2}\frac{(y - \sum_{k=1}^{K} x_k \mu_k)^2}{\sum_{k=1}^{K} x_k \sigma_k^2}\right] \cdot p_1^{x_1} \cdot \ldots \cdot p_K^{x_K}$$

support contains only the values
$(1,0,0,\dots,0)$
$(0,1,0,\dots,0)$
$(0,0,1,\dots,0)$
$\vdots$
$(0,0,0,\dots,1)$

$$= \sum_{k=1}^{K} p_k \cdot \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_k} \exp\left[-\frac{1}{2}\frac{(y-\mu_k)^2}{\sigma_k^2}\right].$$

$\boxed{\text{MULTINOMIAL DISTRIBUTION}}$

Just as we build a Binomial(n, p) r.v. by defining it as the number of successes in n independent Bernoulli trials, we will build what we call Multinomial r.v.s by defining them as the number of times each outcome occurred in n independent Multinoulli trials:

For $k = 1, \dots, K$, let Y_k = # times outcome k occurred in n independent Multinoulli trials with outcome probabilities $p_1, \dots, p_K$.

Then $(Y_1, \dots, Y_K)$ has the Multinomial$(n, p_1, \dots, p_K)$ distribution, which has pmf

$$p(y_1,\dots,y_K) = \left(\frac{n!}{y_1! \cdot \ldots \cdot y_K!}\right) \cdot p_1^{y_1} \cdot \ldots \cdot p_K^{y_K} \cdot \mathbb{1}\Big(\underbrace{(y_1,\dots,y_K) \in \{0,1,\dots,n\}^K,\ \sum_{k=1}^{K} y_k = n}_{\text{the support of } (Y_1,\dots,Y_K)}\Big)$$

Remarks: (i) $\dfrac{n!}{y_1! \cdot \ldots \cdot y_K!}$ = # ways to partition n things into K groups of sizes $y_1, \ldots, y_K$ such that $y_1 + \ldots + y_K = n$.

(ii) For $k = 1, \ldots, K$, the marginal distribution of Y_k is $\text{Binomial}(n, p_k)$.

E.g. Each of 10 customers redeems a coupon for

(1) ice cream with probability $1/4$,
(2) French fries with probability $2/4$,
(3) a donut with probability $1/4$,

and the customers decisions are independent.

Let Y_1 = # out of 10 customers to redeem coupon for ice cream
Y_2 = " French fries
Y_3 = " a donut

(i) Find $P(Y_1 = 5, Y_2 = 3, Y_3 = 2)$:

$$P(Y_1 = 5, Y_2 = 3, Y_3 = 2) = \left(\frac{10!}{5!\,3!\,2!}\right)\left(\tfrac{1}{4}\right)^5\left(\tfrac{1}{2}\right)^3\left(\tfrac{1}{4}\right)^2 = .019$$

in R: dmultinom(x= c(5,3,2) , prob= c(.25,.5,.25))

(ii) Find $P(Y_2 \geq 4)$:

We know that $Y_2 \sim \text{Binomial}\left(10, \tfrac{1}{2}\right)$, so

$$P(Y_2 \geq 4) = \sum_{y_2=4}^{10} \binom{10}{y_2}\left(\tfrac{1}{2}\right)^{y_2}\left(1-\tfrac{1}{2}\right)^{10-y_2} = .828.$$

In R: 1- pbinom(3, 10, .5)

(iii) Find $P(Y_1 = 0 \mid Y_2 = 5)$:

$$P(Y_1 = 0 \mid Y_2 = 5) = \frac{P(Y_1 = 0, Y_2 = 5)}{P(Y_2 = 5)}$$

$$= \frac{P(Y_1 = 0, Y_2 = 5, Y_3 = 5)}{P(Y_2 = 5)}$$

dmultinom(x=c(0,5,5),prob=c(.25,.5,.25))

$$= \frac{\left(\frac{10!}{0!\,5!\,5!}\right)\left(\frac{1}{4}\right)^0\left(\frac{1}{2}\right)^5\left(\frac{1}{4}\right)^5}{\binom{10}{5}\left(\frac{1}{2}\right)^5\left(1-\frac{1}{2}\right)^5} = .0315$$

dbinom(5, 10, .5)

ADDENDUM: Algebraic details for Normal random effects model.

$$\frac{(y-a)^2}{\sigma^2} + \frac{(a-\mu_A)^2}{\sigma_A^2} = \frac{\sigma_A^2\left(y^2 - 2ya + a^2\right) + \sigma^2\left(a^2 - 2a\mu_A + \mu_A^2\right)}{\sigma^2\sigma_A^2}$$

$$= \frac{(\sigma_A^2+\sigma^2)a^2 - 2a(\sigma_A^2 y + \sigma^2\mu_A) + \sigma_A^2 y^2 + \sigma^2\mu_A^2}{\sigma^2\sigma_A^2}$$

$$= \frac{a^2 - 2a\left(\frac{\sigma_A^2 y + \sigma^2\mu_A}{\sigma_A^2+\sigma^2}\right) + \left(\frac{\sigma_A^2 y + \sigma^2\mu_A}{\sigma_A^2+\sigma^2}\right)^2 - \left(\frac{\sigma_A^2 y + \sigma^2\mu_A}{\sigma_A^2+\sigma^2}\right)^2 + \frac{\sigma_A^2 y^2 + \sigma^2\mu_A^2}{(\sigma_A^2+\sigma^2)}}{\sigma^2\sigma_A^2/(\sigma_A^2+\sigma^2)}$$

$$= \frac{\left[a - \left(\frac{\sigma_A^2 y - \sigma^2\mu_A}{\sigma_A^2+\sigma^2}\right)\right]^2}{\sigma^2\sigma_A^2/(\sigma_A^2+\sigma^2)} - \frac{\frac{1}{\sigma^2\sigma_A^2}\left(\sigma_A^4 y^2 + 2\sigma_A^2\sigma^2 y\mu_A + \sigma^4\mu_A^2\right)}{\sigma_A^2+\sigma^2}$$

$$+ \frac{\frac{(\sigma_A^2+\sigma^2)}{\sigma_A^2\sigma^2}\left(\sigma_A^2 y^2 + \sigma^2\mu_A\right)}{\sigma_A^2+\sigma^2}$$

$$= \frac{\left[a - \left(\frac{\sigma_A^2 y - \sigma^2\mu_A}{\sigma_A^2+\sigma^2}\right)\right]^2}{\sigma^2\sigma_A^2/(\sigma_A^2+\sigma^2)} - \frac{\left[\frac{\sigma_A^2}{\sigma^2} y^2 + 2y\mu_A + \frac{\sigma^2}{\sigma_A^2}\mu_A^2\right]}{\sigma_A^2+\sigma^2}$$

$$+ \frac{\left[\left(\frac{\sigma_A^2}{\sigma^2} + 1\right)y^2 + \left(1 + \frac{\sigma^2}{\sigma_A^2}\right)\mu_A\right]}{\sigma_A^2+\sigma^2}$$

$$= \frac{\left[a - \left(\frac{\sigma_A^2 y - \sigma^2\mu_A}{\sigma_A^2+\sigma^2}\right)\right]^2}{\sigma^2\sigma_A^2/(\sigma_A^2+\sigma^2)} + \frac{(y-\mu_A)^2}{\sigma_A^2+\sigma^2}.$$