Two Loaded Dice

$x \backslash y$	1	2	...	6
1	p_{11}	p_{12}	...	p_{16}
2	p_{21}	p_{22}	...	p_{26}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
6	p_{61}	p_{62}	...	p_{66}

$$\mathbb{P}(X+Y=4) = f(1,3) + f(2,2) + f(3,1) = \sum_{x+y=4} f(x,y)$$

$$\mathbb{P}(Y=4) = f(1,4) + f(2,4) + ... + f(6,4) = \sum_{x} f(x,4)$$

Two Random Variables

Joint Distribution $f(x, y)$

$$\mathbb{P}[g(X,Y) \in A] \quad = \quad \sum_{g(x,y)\in A} f(x,y) \quad \text{or} \quad \int_{g(x,y)\in A} f(x,y)\,dx\,dy$$

Marginal Distribution $f_X(x), f_Y(y)$

$$f_X(x) = \sum_{y} f(x,y) \text{ or } \int f(x,y)\,dy \quad [\text{likewise for } f_Y(y)]$$

Conditional Distribution $f_{Y|X}(y|x), f_{X|Y}(x|y)$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \quad [\text{likewise for } f_{X|Y}(x|y)]$$

Exercise 3.1 (p. 32)

Question Suppose X and Y are independent random variables, each distributed uniformly on the interval $(0, 1)$. What's $\mathbb{P}(X - Y > 1/4)$?

Context Amy and Bob agree to meet at the library to work on homework together. Independently, they are both equally likely to arrive at any time between 7 pm and 8 pm. What's the probability that Amy has to wait at least 15 minutes for Bob?

Example 3.3 + Exercise 3.2 (pp. 32–34)

Trinomial$(n; p_1, p_2, p_3)$

$$f(x,y) = \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1-p_1-p_2)^{n-x-y}, \quad x, y \in \{0, 1, ..., n\}$$

Remark Special case of multinomial$(n; p_1, p_2, ..., p_K)$,

$$f(x_1, x_2, ..., x_K) = \frac{n!}{x_1!x_2!...x_K!} p_1^{x_1} p_2^{x_2} ... p_K^{x_K}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} p_k = 1 \quad \text{and} \quad \sum_{k=1}^{K} x_k = n.$$

If $K = 2$, get the binomial(n, p).

Marginal

$$(a+b)^m = \sum_{i=0}^{m} \frac{m!}{i!(m-i)!} a^i b^{m-i}$$

$$f_X(x) = \sum_{y=0}^{??} \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1-p_1-p_2)^{n-x-y}$$

$$\vdots$$

$$= \frac{n!}{x!(n-x)!} p_1^x (1-p_1)^{n-x}$$

$$\sim \text{binomial}(n, p_1)$$

Conditional

$$f_{Y|X}(y|x) = \frac{\left[\frac{n!}{x!y!(n-x-y)!}p_1^x p_2^y (1-p_1-p_2)^{n-x-y}\right]}{\left[\frac{n!}{x!(n-x)!}p_1^x (1-p_1)^{n-x}\right]}$$

$$\vdots$$

$$= \frac{(n-x)!}{y!(n-x-y)!}\left[\frac{p_2}{1-p_1}\right]^y \left[1-\frac{p_2}{1-p_1}\right]^{n-x-y}$$

$$\sim \text{binomial}\left(n-x; \frac{p_2}{1-p_1}\right)$$

Remark Conclusions about marginal and conditional generalizes to $K > 3$.

Example 3.4 (pp. 34–35)

Joint

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right], \quad x, y \in \mathbb{R}$$

Remark Special case of n-dimensional normal$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

$$f(\boldsymbol{z}) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp\left[-\frac{(\boldsymbol{z}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}-\boldsymbol{\mu})}{2}\right], \quad \boldsymbol{z} \in \mathbb{R}^n,$$

with

$$\boldsymbol{z} = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2, \quad \boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Example 3.4 (pp. 34–35)

Marginal

$$
\begin{aligned}
f_X(x) &= \int f(x,y)dy \\
&= \int \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{y^2 - 2\rho xy + \rho^2 x^2 + (1-\rho^2)x^2}{2(1-\rho^2)}\right] dy \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \times \underbrace{\int \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right] dy}_{\mathrm{N}(\rho x, 1-\rho^2)} \\
&\sim \mathrm{N}(0,1)
\end{aligned}
$$

Conditional

$$
\therefore \quad Y|(X=x) \sim \mathrm{N}(\rho x, 1-\rho^2) \qquad [\text{why}]
$$

Conditional Expectations/Variances

Recall ...

$$\mathbb{E}(Y) = \sum_y y f(y) \ \underline{\text{or}} \ \int_{\mathbb{R}} y f(y) dy, \quad \mathbb{V}\text{ar}(Y) = \mathbb{E}\{[Y - \mathbb{E}(Y)]^2\}$$

Definition

$$\mathbb{E}(Y|X=x) = \sum_y y f_{Y|X}(y|x) \ \underline{\text{or}} \ \int_{\mathbb{R}} y f_{Y|X}(y|x) dy$$

$$\mathbb{V}\text{ar}(Y|X=x) = \mathbb{E}\{[Y - \mathbb{E}(Y|X=x)]^2 | X=x\}$$

Remark The textbook, *Essential Statistics*, does <u>not</u> really cover these concepts.

A Trivial Example

$f(x, y)$

$x \backslash y$	0	1
0	0.5	0
1	0	0.5

(a) Distribution of Y, $\mathbb{E}(Y)$, and $\mathbb{V}\text{ar}(Y)$.

(b) Distribution of $Y|X = 1$, $\mathbb{E}(Y|X = 1)$, and $\mathbb{V}\text{ar}(Y|X = 1)$.

(c) Distribution of $Y|X = 0$, $\mathbb{E}(Y|X = 0)$, and $\mathbb{V}\text{ar}(Y|X = 0)$.

Earlier Examples Today

Trinomial $Y|(X=x) \sim \text{binomial}[n-x, p_2/(1-p_1)]$, so

$$\begin{aligned} \mathbb{E}(Y|X) &= (n-X)\left(\frac{p_2}{1-p_1}\right) \\ \mathbb{V}\text{ar}(Y|X) &= (n-X)\left(\frac{p_2}{1-p_1}\right)\left(1-\frac{p_2}{1-p_1}\right) \end{aligned}$$

Special Bivariate Normal $Y|(X=x) \sim \text{N}(\rho x, 1-\rho^2)$, so

$\mathbb{E}(Y|X) = \rho X$ linear function of X

$\mathbb{V}\text{ar}(Y|X) = 1-\rho^2$ independent of X

More Than Two RVs

- same ideas, "just" more tedious — mostly, high-dimensional sums and integrals, e.g.,

$$\sum_{g(x_1,...,x_n)\in A} f(x_1,...,x_n) \quad \text{or} \quad \int_{g(x_1,...,x_n)\in A} f(x_1,...,x_n)dx_1...dx_n$$

- often partition into two blocks, e.g., for $\boldsymbol{Z} \in \mathbb{R}^{1000}$,

$$\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_{800} \\ \hline Z_{801} \\ \vdots \\ Z_{1000} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X} \\ \hline \boldsymbol{Y} \end{bmatrix}, \quad \boldsymbol{X} \in \mathbb{R}^{800}, \boldsymbol{Y} \in \mathbb{R}^{200},$$

and really think as if just two RVs — $f(\boldsymbol{x}, \boldsymbol{y}), f(\boldsymbol{x}), f(\boldsymbol{y}|\boldsymbol{x})$

Multivariate Normal in $\mathbb{R}^n$

- if $\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and

$$\boldsymbol{Z} = \left[\begin{array}{c} \boldsymbol{X} \\ \hline \boldsymbol{Y} \end{array}\right] \in \mathbb{R}^n, \quad \boldsymbol{\mu} = \left[\begin{array}{c} \boldsymbol{\mu}_1 \\ \hline \boldsymbol{\mu}_2 \end{array}\right], \quad \boldsymbol{\Sigma} = \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right],$$

  then the same technique can be applied to derive that

$$\boldsymbol{Y}|\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$$

- notice $\mathbb{E}(\boldsymbol{Y}|\boldsymbol{X})$ is a linear function, and $\mathbb{V}\mathrm{ar}(\boldsymbol{Y}|\boldsymbol{X})$ is independent, of $\boldsymbol{X}$

Multivariate Normal and Classification

$$f(z) = \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}|}} \exp\left[-\frac{(\boldsymbol{z}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})}{2}\right]$$

- $f_1 \sim \mathrm{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$, $f_2 \sim \mathrm{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$: linear decision boundary

  want $\quad f_1(\boldsymbol{z}) = f_2(\boldsymbol{z}) \quad$ (recall Lecture 3 + Exercise 2.11)

$$(\boldsymbol{z}-\boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1) = (\boldsymbol{z}-\boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_2)$$

$$\boldsymbol{z}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{z} - 2\boldsymbol{\mu}_1^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{z} + \boldsymbol{\mu}_1^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 = \boldsymbol{z}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{z} - 2\boldsymbol{\mu}_2^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{z} + \boldsymbol{\mu}_2^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2$$

$$\Rightarrow \quad \underbrace{2(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}^{-1}}_{\boldsymbol{w}^\top}\boldsymbol{z} + \underbrace{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)}_{w_0} = 0$$

- $f_1 \sim \mathrm{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $f_2 \sim \mathrm{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$: quadratic decision boundary