Chapter 2. Probability

2.1.
a. 2/3. The sample space is $\{BB, BG, GB, GG\}$. Since the neighbor has any boy, the case GG is ruled out. Therefore, the probability that the second child is a girl is 2/3.

b. 1/2. The only uncertainty remaining is the gender of the child unseen, and given no any other information, the probability would be 1/2.

2.2.
a. Denote the null hypothesis that the defendant is innocent be I, the alternative hypothesis that the defendant is guilty be G, and the evidence of the blood type found be E. We know that $p(E|I) = 1/100$. The prosecutor states that $p(G|E) = 1 - p(I|E) = 1 - p(E|I) = 99/100$, which is not true, because $p(I|E) \neq p(E|I)$.

b. The accused was not randomly chosed from all people of the city having his blood type, and has additional evidences to the case. That other informations needs to be taken into account. If no informations other than the blood type are given, then the defender's argument would make sense.

2.3.
$\mathrm{Var}[X+Y] = \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2 = \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X]^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y]^2) = (\mathbb{E}[X^2] - \mathbb{E}[X]^2) + (\mathbb{E}[Y^2] - \mathbb{E}[Y]^2) + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y]$

2.4.
$p(P|D) = p(N|\bar{D}) = 0.99$, $p(D) = 1/10000$.

$$p(D|P) = \frac{p(D)p(P|D)}{p(D)p(P|D) + p(\bar{D})p(P|\bar{D})} = \frac{\frac{1}{10000} \times \frac{99}{100}}{\frac{1}{10000} \times \frac{99}{100} + \frac{9999}{10000} \times \frac{1}{100}} \approx 0.0098$$

2.5
The contestant should switch to another door.

$$p(C_1|D_3) = \frac{p(C_1)p(D_3|C_1)}{p(C_1)p(D_3|C_1) + p(C_2)p(D_3|C_2) + p(C_3)p(D_3|C_3)} = \frac{\frac{1}{3}\cdot\frac{1}{2}}{\frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot 0} = \frac{1}{3},$$

$$p(C_2|D_3) = \frac{p(C_2)p(D_3|C_2)}{p(C_1)p(D_3|C_1) + p(C_2)p(D_3|C_2) + p(C_3)p(D_3|C_3)} = \frac{\frac{1}{3}\cdot 1}{\frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot 0} = \frac{2}{3}.$$

2.6
a.

$$P(H|e_1, e_2) = \frac{P(e_1, e_2|H)P(H)}{P(e_1, e_2)},$$

so the set ii. is required.
b.

$$\begin{aligned} P(H|e_1, e_2) = \frac{P(e_1, e_2, H)}{P(e_1, e_2)} &= \frac{P(e_1|e_2, H)P(e_2, H)}{P(e_1, e_2)} \\ &= \frac{P(e_1|H)P(e_2, H)}{P(e_1, e_2)} = \frac{P(e_1|H)P(e_2|H)P(H)}{P(e_1, e_2)} \end{aligned} \tag{1}$$

so the set i. is required.

2.7
Suppose X and Y are two independent tosses of a fair coin. Let $Z = 1$ if exactly one of X, Y is 1, and 0 otherwise. Then X, Y, Z are pairwise independent, but not mutually independent, since $p(X = 0, Y = 0, Z = 0) = 1/4$ but $p(X = 0)p(Y = 0)p(Z = 0) = 1/8$.

2.8
$\Rightarrow$:

$$p(x, y|z) = p(x|z)p(y|z) = \frac{p(x, z)}{p(z)}\frac{p(y, z)}{p(z)} = g(x, z)h(y, z).$$

$\Leftarrow$:

$$p(x|z) = \int p(x, y|z)dy = g(x, z)\int h(y, z)dy.$$

Therefore, $g(x,z) = \alpha(z)p(x|z)$ for some $\alpha(z) > 0$. Symmetrically, we can derive $h(y,z) = \beta(z)p(y|z)$ for some $\beta(z) > 0$. Plugging these results gives $p(x,y|z) = \alpha(z)\beta(z)p(x|z)p(y|z)$. Since

$$\int\int p(x,y|z)dxdy = \alpha(z)\beta(z)\int p(x|z)dx\int p(y|z)dy = 1,$$

we have $\alpha(z)\beta(z) = 1$. Hence $p(x,y|z) = p(x|z)p(y|z)$.

2.9
a. True. We have $X \perp W|Z,Y \Rightarrow p(X,W|Z,Y) = p(X|Z,Y)p(W|Z,Y)$

$$\Rightarrow p(X,W,Z,Y) = \frac{p(X,Z,Y)p(W,Z,Y)}{p(Z,Y)}$$

$X \perp Y|Z \Rightarrow p(X,Y|Z) = p(X|Z)p(Y|Z)$

$$\Rightarrow p(X,Y,Z) = \frac{p(X,Z)p(Y,Z)}{p(Z)}$$

To show $X \perp Y,W|Z$, it suffices to verify whether $p(X,Y,W|Z) = p(X|Z)p(Y,W|Z)$.

$$\frac{p(X,Y,W|Z)}{p(X|Z)p(Y,W|Z)} = \frac{p(X,Y,W,Z)p(Z)}{p(X,Z)p(Y,W,Z)} = \frac{p(Y,W,Z)p(X,Y,Z)p(Z)}{p(Y,Z)p(X,Z)p(Y,W,Z)}$$

$$= \frac{p(X,Y,Z)p(Z)}{p(Y,Z)p(X,Z)} = \frac{p(X,Z)p(Y,Z)p(Z)}{p(Z)p(Y,Z)p(X,Z)} = 1.$$

Therefore $X \perp Y,W|Z$.

b. False. Consider three i.i.d r.v. X,Y,Z having the values of -1 or 1 equally likely. Let $W = XYZ$. Then clearly $X \perp Y|Z$ and $X \perp Y|W$, but $X \perp Y|Z,W$ does not hold, since $ZW = XY$.

2.10

$$p_y(y) = p_x(x)|\frac{dx}{dy}| = \frac{b^a}{\Gamma(a)}y^{1-a}e^{-\frac{b}{y}}\frac{1}{y^2} = \frac{b^a}{\Gamma(a)}y^{-(a+1)}e^{-\frac{b}{y}}.$$

2.11
With using the change of variables $u = e^{-\frac{r^2}{2\sigma^2}}$,

$$z^2 = \int_0^{2\pi} d\theta \int_0^{\infty} re^{-\frac{r^2}{2\sigma^2}}dr = (2\pi)\sigma^2\int_0^1 du = 2\pi\sigma^2.$$

2.12

$$
\begin{aligned}
\mathbb{I}(X,Y) &= \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\
&= \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)} - \sum_x \sum_y p(x,y) \log p(y) \\
&= \sum_x \sum_y p(x)p(y|x) \log p(y|x) - \sum_x \sum_y p(x,y) \log p(y) \\
&= \sum_x (p(x)(\sum_y p(y|x) \log p(y|x))) - \sum_y \log p(y) \sum_x p(x,y) \\
&= -\sum_x p(x)\mathbb{H}(Y|X=x) - \sum_y \log p(y) \log p(x) \\
&= -\mathbb{H}(Y|X) + \mathbb{H}(Y) = \mathbb{H}(Y) - \mathbb{H}(Y|X).
\end{aligned} \tag{2}
$$

The property $\mathbb{I}(X,Y) = \mathbb{H}(X) - \mathbb{H}(X|Y)$ can be derived similarly.

2.13

$$
f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)\sigma^2}(x_1^2 - 2\rho x_1 x_2 + x_2^2)}
$$

The marginal distribution of X_1 is

$$
f_{X_1}(x_1) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-\frac{x_1^2}{2\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x_2-\rho x_1)^2}{2(1-\rho^2)\sigma^2}} dx_2 = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x_1^2}{2\sigma^2}}.
$$

Similarly,

$$
f_{X_2}(x_2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x_2^2}{2\sigma^2}}.
$$

Now the entropy for X_1 is

$$
\mathbb{H}(X_1) = -\int_{-\infty}^{\infty} \log f_{X_1}(x_1) f_{X_1}(x_1) dx_1 = -\int_{-\infty}^{\infty} (\log(\frac{1}{\sqrt{2\pi}\sigma}) - \frac{x_1^2}{2\sigma^2}) f_{X_1}(x_1) dx_1
$$

$$
= -\log(\frac{1}{\sqrt{2\pi}\sigma}) \cdot 1 + \frac{1}{2\sigma^2} \mathrm{Var}[\mathcal{N}(0,\sigma^2)] = \frac{1}{2}\log(2\pi e\sigma^2).
$$

Similarly, $\mathbb{H}(X_2) = \frac{1}{2}\log(2\pi e\sigma^2)$.
To go further, we must check the following property:

$$\mathbb{H}(X_1, X_2) = \mathbb{H}(X_1) + \mathbb{H}(X_2|X_1).$$

To show this,

$$\begin{aligned}
\mathbb{H}(X_1, X_2) &= -\sum_{x_1}\sum_{x_2} p(x_1, x_2)\log p(x_1, x_2) = -\sum_{x_1}\sum_{x_2} p(x_1, x_2)\log(p(x_1)p(x_2|x_1)) \\
&= -\sum_{x_1}\sum_{x_2} p(x_1, x_2)\log p(x_1) - \sum_{x_1}\sum_{x_2} p(x_1, x_2)\log p(x_2|x_1) \\
&= -\sum_{x_1}\log p(x_1)(\sum_{x_2} p(x_1, x_2)) - \sum_{x_1} p(x_1)(\sum_{x_2} p(x_2|x_1)\log p(x_2|x_1)) \\
&= -\sum_{x_1}\log p(x_1)p(x_1) - \sum_{x_1} p(x_1)\mathbb{H}(X_2|X_1 = x_1) = \mathbb{H}(X_1) + \mathbb{H}(X_2|X_1).
\end{aligned} \tag{3}$$

Using this lemma, we have $\mathbb{I}(X_1, X_2) = \mathbb{H}(X_1) + \mathbb{H}(X_2) + \mathbb{H}(X_1, X_2)$.
To compute $\mathbb{H}(X_1, X_2) = \mathbb{H}(\mathbf{X})$,

$$\mathbb{H}(\mathbf{X}) = -\int_{\mathbf{x}} \log f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = -\int_{\mathbf{x}} (\log(\frac{1}{\sqrt{\det(2\pi\mathbf{\Sigma})}}) - \frac{(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

$$= -\log(\frac{1}{\sqrt{\det(2\pi\mathbf{\Sigma})}}) + \frac{1}{2}\mathbb{E}[(\mathbf{X}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})].$$

To compute the second term, since the inner expression is scalar,

$$\mathbb{E}[(\mathbf{X}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})] = \mathbb{E}[\mathrm{trace}((\mathbf{X}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}))] = \mathbb{E}[\mathrm{trace}(\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})^{\mathrm{T}}(\mathbf{X}-\boldsymbol{\mu}))]$$

$$= \mathrm{trace}(\mathbf{\Sigma}^{-1}\mathbb{E}[(\mathbf{X}-\boldsymbol{\mu})^{\mathrm{T}}(\mathbf{X}-\boldsymbol{\mu})]) = \mathrm{trace}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}) = \dim(\mathbf{X}).$$

Therefore,

$$\mathbb{H}(\mathbf{X}) = -\log(\frac{1}{\sqrt{\det(2\pi\mathbf{\Sigma})}}) + \frac{1}{2}\dim(\mathbf{X}) = \frac{1}{2}\log(\det(2\pi e\mathbf{\Sigma})).$$

In the problem, $\dim(\mathbf{X}) = 2$ and $\det(\mathbf{\Sigma}) = \sigma^4(1-\rho^2)$, thus
$\mathbb{H}(X_1, X_2) = \frac{1}{2}\log(4\pi^2\sigma^4(1-\rho^2))$.

$$\Rightarrow \mathbb{I}(X_1, X_2) = \mathbb{H}(X_1) + \mathbb{H}(X_2) - \mathbb{H}(X_1, X_2) = \frac{1}{2}\log(\frac{1}{1-\rho^2}).$$

If $\rho = \pm 1$, $\mathbb{I}(X_1, X_2) = -\infty$, meaning that X_1, X_2 are fully correlated.
If $\rho = 0$, $\mathbb{I}(X_1, X_2) = 0$, meaning that X_1, X_2 are independent.

2.14
a. $\mathbb{I}(X, Y) = \mathbb{H}(X) - \mathbb{H}(X|Y) \Rightarrow r = \frac{\mathbb{I}(X,Y)}{\mathbb{H}(X)}$.

b. Since $\mathbb{I}(X, Y) \geq 0$ and $\mathbb{H}(X) \geq 0$, $r \geq 0$.
Since $\mathbb{H}(X|Y) \geq 0$, $r \leq 1$.

c. $r = 0 \Rightarrow \mathbb{H}(X|Y) = \mathbb{H}(X)$.
X, Y are independent and identically distributed.

d. $r = 1 \Rightarrow \mathbb{H}(X|Y) = 0$.
Y is completely determined by X.

2.15
We have that $\mathbb{KL}(p_{emp}||q) = 0$ if and only if $p_{emp} = q$.
If $p_{emp} = q$, the log likelihood $\sum_i \log(q(x_i; \boldsymbol{\theta}))$ is maximized, so $p_{emp} = q(x, \hat{\boldsymbol{\theta}})$.

2.16
Mean:

$$\int_0^1 x \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} dx = \frac{B(a+1)}{B(a,b)} = \frac{a}{a+b}.$$

Variance:

$$\int_0^1 x^2 \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} dx = \frac{B(a+2)}{B(a,b)} = \frac{a(a+1)}{(a+b)(a+b+1)}.$$

Mode:

$$\frac{d}{dx}\mathrm{Beta}(x|a,b) = 0$$

$$\Rightarrow x^{a-2}(1-x)^{b-2}((a-1)x - (b-1)(1-x)) = 0$$

$$\Rightarrow x = \frac{a-1}{a+b-2}.$$

2.17

Let $Z = \min(X, Y)$.

$P(Z \leq z) = 1 - P(Z \geq z) = 1 - P(X \geq z)P(Y \geq z) = 1 - (1 - z)^2$.

Therefore, $f_Z(z) = 2z - 2$.

$$\mathbb{E}[Z] = \int_0^1 z(2z - 2)dz = \frac{1}{3}.$$