ECE 361 Probability for Engineers (Fall, 2016)
Lecture 7a

§3.5 Conditioning

Conditioning an RV on an event

The total probability theorem for conditional PDFs states: given a collection of events $A_1, \ldots, A_n$ that partition the sample space with $\mathbb{P}(A_i) > 0$ for all i then

$$f_X(x) = \sum_{i=1}^{n} \mathbb{P}(A_i) f_{X|A_i}(x). \tag{1}$$

Example. The metro train arrives every quarter hour. You arrive at the train station uniformly between 7:10 and 7:30am. What is the PDF of the time you wait to board the train? Let $X \sim \text{Uni}[10, 30]$ be the time of our arrival at the train station and Y be the waiting time to board the train. Note $Y \in [0, 15]$. Define $A = \{X \in [10, 15]\}$ and $B = \{X \in [15, 30]\}$. Conditioned on A, Y is uniform over 0 and 5 while conditioned on B, Y is uniform over $[0, 15]$. Thus:

$$f_Y(y) = \mathbb{P}(A) f_{Y|A}(y) + \mathbb{P}(B) f_{Y|B}(y) = \left\{ \begin{array}{ll} \frac{1}{4}\frac{1}{5} + \frac{3}{4}\frac{1}{15} = \frac{1}{10}, & y \in [0, 5] \\ \frac{1}{4}0 + \frac{3}{4}\frac{1}{15} = \frac{1}{20}, & y \in (5, 15] \end{array} \right. \tag{2}$$

Conditioning one RV on another

Consider a pair of RVs (X, Y) and a y with $f_Y(y) > 0$. Then the conditional PDF of X given $Y = y$ is:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}. \tag{3}$$

Note that the marginalization equation $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\mathrm{d}x$ implies normalization for the conditional PDF:

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)\mathrm{d}x = \int_{-\infty}^{\infty} \frac{f_{X,Y}(x, y)}{f_Y(y)}\mathrm{d}x = 1. \tag{4}$$

Circuluar uniform PDF. Ben throws a dart at a circle of radius r – suppose all points are equally likely so (X, Y) is uniform over the circle:

$$f_{X,Y}(x, y) = \left\{ \begin{array}{ll} \frac{1}{\pi r^2}, & x^2 + y^2 \leq r^2 \\ 0, & \text{else} \end{array} \right. \tag{5}$$

Find the conditional PDF $f_{X|Y}(x|y)$ by first finding the marginal PDF $f_Y(y)$:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\mathrm{d}x = \frac{1}{\pi r^2} \int_{(x,y):x^2+y^2\leq r^2} \mathrm{d}x = \frac{1}{\pi r^2} \int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} \mathrm{d}x = \frac{2}{\pi r^2}\sqrt{r^2 - y^2}, \ \forall y : |y| \leq r. \tag{6}$$

Now the conditional PDF is:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{\frac{1}{\pi r^2}}{\frac{2}{\pi r^2}\sqrt{r^2 - y^2}} = \frac{1}{2\sqrt{r^2 - y^2}}, \ \forall (x, y) : x^2 + y^2 \leq r^2. \tag{7}$$

To interpret the conditional PDF, fix small (δ_1, δ_2) and values (x, y) and consider

$$\mathbb{P}(X \in [x, x + \delta_1]|Y \in [y, y + \delta_2]) = \frac{\mathbb{P}(X \in [x, x + \delta_1], Y \in [y, y + \delta_2])}{\mathbb{P}(Y \in [y, y + \delta_2])} \approx \frac{f_{X,Y}(x, y)\delta_1\delta_2}{f_Y(y)\delta_2} = f_{X|Y}(x|y)\delta_1. \tag{8}$$

Thus:

$$\mathbb{P}(X \in [x, x + \delta]|Y = y) \approx f_{X|Y}(x|y)\delta_1, \tag{9}$$

or $f_{X|Y}(x|y)$ is the probability per unit length for X around x given $Y = y$.

Example. A vehicle's speed X is exponential with mean $\lambda = 50$ mph. Suppose the police radar measurement Y has a normally distributed random error with zero mean and standard deviation of one tenth of the speed of the vehicle. What is the joint PDF for X, Y? First:

$$f_X(x) = \frac{1}{50}\mathrm{e}^{-x/50}. \tag{10}$$

Next, conditioned on $X = x$ we have that $Y = x + N(0, x/10) \sim N(x, x/10)$:

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}(x/10)}\mathrm{e}^{-\frac{(y-x)^2}{2x^2/100}}. \tag{11}$$

Then:

$$f_{X,Y}(x,y) = f_X(x)f_{Y|X}(y|x) = \frac{1}{50}\mathrm{e}^{-x/50}\frac{1}{\sqrt{2\pi}(x/10)}\mathrm{e}^{-\frac{(y-x)^2}{2x^2/100}}. \tag{12}$$

Conditional expectation

The conditional expectation of an RV X conditioned on an event A is defined as:

$$\mathbb{E}[X|A] = \int_{-\infty}^{\infty} x f_{X|A}(x)\mathrm{d}x. \tag{13}$$

For an event $A = \{Y = y\}$ is

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\mathrm{d}x. \tag{14}$$

For a function $g(X)$:

$$\begin{aligned} \mathbb{E}[g(X)|A] &= \int_{-\infty}^{\infty} g(x) f_{X|A}(x)\mathrm{d}x \\ \mathbb{E}[g(X)|Y = y] &= \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\mathrm{d}x. \end{aligned} \tag{15}$$

For a partition $(A_1, \ldots, A_n)$ with $\mathbb{P}(A_i) > 0$ for each $i \in [n]$:

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{E}[X|A_i]. \tag{16}$$

Similarly, we can condition on all possible values of an RV Y:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X|Y = y] f_Y(y)\mathrm{d}y. \tag{17}$$

Analogously for $g(X)$:

$$\begin{aligned} \mathbb{E}[g(X,Y)|Y = y] &= \int g(x,y) f_{X|Y}(x|y)\mathrm{d}x \\ \mathbb{E}[g(X,Y)] &= \int \mathbb{E}[g(X,Y)|Y = y] f_Y(y)\mathrm{d}y. \end{aligned} \tag{18}$$

Example. (Mean and variance of a piecewise constant PDF.) Suppose X is piecewise constant:

$$f_X(x) = \left\{ \begin{array}{ll} 1/3, & x \in [0,1] \\ 2/3, & x \in (1,2] \end{array} \right. \tag{19}$$

Consider events $A_1 = \{X \in [0,1]\}$ and $A_2 = \{X \in (1,2]\}$. Then:

$$\mathbb{P}(A_1) = 1/3, \ \mathbb{P}(A_2) = 2/3. \tag{20}$$

Further:

$$\mathbb{E}[X|A_1] = 1/2, \ \mathbb{E}[X|A_2] = 3/2 \tag{21}$$

and

$$\begin{aligned}
\mathbb{E}[X^2|A_1] &= \int_0^1 x^2 f_{X|A}(x)\mathrm{d}x = \frac{1}{\mathbb{P}(A_1)} \int_0^1 x^2 f_X(x)\mathrm{d}x = 3\frac{1}{3}\left.\frac{1}{3}x^3\right|_0^1 = \frac{1}{3} \\
\mathbb{E}[X^2|A_2] &= \cdots = \frac{7}{3}
\end{aligned} \tag{22}$$

Then:

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X|A_1]\mathbb{P}(A_1) + \mathbb{E}[X|A_2]\mathbb{P}(A_2) = \frac{1}{2}\frac{1}{3} + \frac{3}{2}\frac{2}{3} = \frac{7}{6} \\
\mathbb{E}[X^2] &= \mathbb{E}[X^2|A_1]\mathbb{P}(A_1) + \mathbb{E}[X^2|A_2]\mathbb{P}(A_2) = \frac{1}{3}\frac{1}{3} + \frac{7}{3}\frac{2}{3} = \frac{15}{9}
\end{aligned} \tag{23}$$

Thus the variance is

$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{15}{9} - \frac{49}{36} = \frac{11}{36}. \tag{24}$$

Independence

Two continuous RVs are independent if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all (x,y) with $f_Y(y) > 0$. Equivalently, $f_{X|Y}(x|y) = f_X(x)$ for all (x,y) with $f_Y(y) > 0$.

Example. (Independent normal RVs.) Let $X \sim N(\mu_x, \sigma_x)$ and $Y \sim N(\mu_y, \sigma_y)$ be independent normal RVs. Their joint PDF is

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}\right\}. \tag{25}$$

See Fig. 3.19 for a visualization of the PDF and its elliptical contours.

For independent RVs (X,Y) and subsets A, B of $\mathbb{R}$ we have

$$\mathbb{P}(X \in A \cap Y \in B) = \int_{x\in A}\int_{y\in B} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = \int_{x\in A} f_X(x)\mathrm{d}x \int_{y\in B} f_Y(y)\mathrm{d}y = \mathbb{P}(X \in A)\mathbb{P}(Y \in B). \tag{26}$$

For $A = \{X \leq x\}$ and $B = \{Y \leq y\}$ we see:

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y) = F_X(x)F_Y(y). \tag{27}$$

It follows that for independent RVs (X,Y) and functions $g(X), h(Y)$:

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]. \tag{28}$$

§3.6 The continuous Bayes' rule

Recall in §1.4 we presented Bayes' rule for events: we think of an event B as the effect (which is observable, i.e., we will know whether the random outcome is in B or B^c), and we are interested in ascertaining the cause (which is not observable), represented by the partition $A_1, \ldots, A_n$ of Ω. We are interested in the probabilities of each possible cause given the effect $\mathbb{P}(A_i|B)$, but our model tells us only the probability of the effect given the cause $\mathbb{P}(B|A_i)$. Bayes' rule allows us to obtain each $\mathbb{P}(A_i|B)$ from all the $\mathbb{P}(B|A_i)$ and $\mathbb{P}(A_i)$ values:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^n \mathbb{P}(B|A_j)\mathbb{P}(A_j)}, \ i \in [n]. \tag{29}$$

Now we address Bayes' rule for continuous RVs; see Fig. 3.20. Here RV X is an unobserved phenomenon characterized by PDF f_X, RV Y is an observation with PDF f_Y related to X by a conditional PDF $f_{Y|X}(y|x)$, and the objective is to infer $f_{X|Y}(x|y)$, the distribution of the phenomenon conditioned on the observation. Bayes's rule in this context is:

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)} = \frac{f_X(x) f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|t) f_X(t) \mathrm{d}t}. \tag{30}$$

Example. A lightbulb is known to have an exponentially distributed lifetime Y with some parameter λ. But due to quality control problems at the factory, λ is itself an RV where $\lambda \sim \mathrm{Uni}[1, 3/2]$. We test a lightbulb and record its lifetime – what can we say about the underlying parameter λ. Here:

$$f_\Lambda(\lambda) = 2, \ 1 \leq \lambda \leq 3/2. \tag{31}$$

Then:

$$f_{\Lambda|Y}(\lambda|y) = \frac{f_\Lambda(\lambda) f_{Y|\Lambda}(y|\lambda)}{\int_{-\infty}^{\infty} f_{Y|\Lambda}(y|t) f_\Lambda(t) \mathrm{d}t} = \frac{2\lambda \mathrm{e}^{-\lambda y}}{\int_1^{3/2} 2t\mathrm{e}^{-ty} \mathrm{d}t} = \frac{2\mathrm{e}^{(3/2-\lambda)y} y^2 \lambda}{-2 - 3y + 2\mathrm{e}^{y/2}(1+y)}. \tag{32}$$

See Fig. 1 (left).

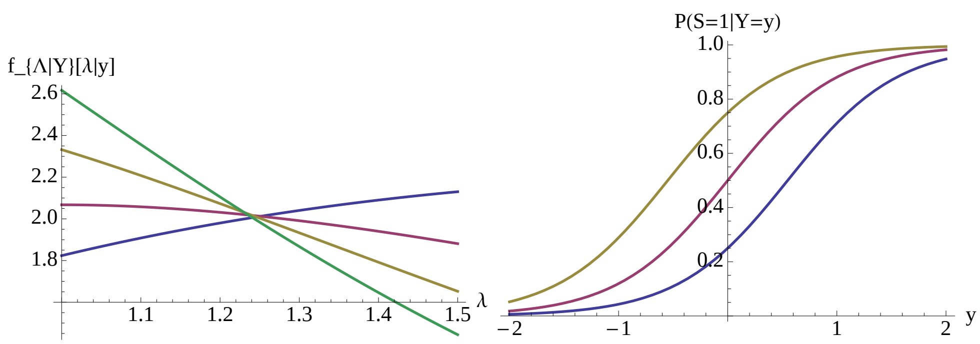

Figure 1: **Left:** the conditional PDF $f_{\Lambda|Y}(\lambda|y)$ for the lightbulb example for $y \in \{1/2, 1, 3/2, 2\}$. The $y = 1/2$ curve is blue and the $y = 2$ curve is green. **Right:** the conditional PMF $\mathbb{P}(S = 1|Y = y)$ for $p \in \{1/4, 1/2, 3/4\}$. The $p = 1/4$ curve is blue and the $p = 3/4$ curve is yellow.

Inference about a discrete RV

Suppose the unobserved phenomenon is either present or absent; let A be the event that the phenomenon is present and $\mathbb{P}(A)$ be its probability of occurrence. Given an observation $Y = y$ we are interested in:

$$\mathbb{P}(A|Y = y) \approx \mathbb{P}(A|y \leq Y \leq y + \delta) = \frac{\mathbb{P}(A)\mathbb{P}(y \leq Y \leq y + \delta|A)}{\mathbb{P}(y \leq Y \leq y + \delta)} = \frac{\mathbb{P}(A) f_{Y|A}(y)\delta}{f_Y(y)\delta} = \frac{\mathbb{P}(A) f_{Y|A}(y)}{f_Y(y)} \tag{33}$$

Now use the TPT on the denominator conditioning on A:

$$\mathbb{P}(A|Y = y) = \frac{\mathbb{P}(A) f_{Y|A}(y)}{\mathbb{P}(A) f_{Y|A}(y) + \mathbb{P}(A^c) f_{Y|A^c}(y)} \tag{34}$$

If the unobserved phenomenon is a discrete RV N then we can use the above for events $A = \{N = n\}$ for each n:

$$\mathbb{P}(N = n|Y = y) = \frac{p_N(n) f_{Y|N}(y|n)}{\sum_i p_N(i) f_{Y|N}(y|i)}. \tag{35}$$

Signal detection. A binary signal $S \in \{-1, 1\}$ is transmitted with $\mathbb{P}(S = 1) = p = 1 - \mathbb{P}(S = -1)$. The signal is corrupted by noise so the observation is $Y = S + N$ where $N \sim N(0, 1)$ and N is independent of S. Note conditioned on $S = s$ we have

$Y \sim N(s, 1)$. We are interested in :

$$\mathbb{P}(S=1|Y=y) = \frac{p_S(1) f_{Y|S}(y|1)}{p_S(1) f_{Y|S}(y|1) + p_S(-1) f_{Y|S}(y|-1)} = \frac{p \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(y-1)^2}{2}}}{p \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(y-1)^2}{2}} + (1-p) \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(y-(-1))^2}{2}}} = \frac{p\mathrm{e}^{y}}{p\mathrm{e}^{y} + (1-p)\mathrm{e}^{-y}}. \tag{36}$$

See Fig. 1 (right).

References

[1] *Introduction to Probability, 2nd Edition* by Dimitri P. Bertsekas and John N. Tsitsiklis, Athina Scientific Press, 2008.