Lecture 7a

ECE 361
Probability for Engineers
Fall, 2016
Steven Weber

Tuesday November 3, 2016

DREXEL UNIVERSITY
Electrical and Computer Engineering
College of Engineering

Outline

1. §3.5 Conditioning
   - Conditioning an RV on an event
   - Conditioning one RV on another
   - Conditional expectation
   - Independence

2. §3.6 The continuous Bayes' rule
   - Inference about a discrete RV

Outline

Conditioning an RV on an event

Fix a collection of events $A_1, \ldots, A_n$ that partition the sample space. The total probability theorem for a discrete RV X states:

$$p_X(x) = \sum_{i=1}^{n} \mathbb{P}(A_i) p_{X|A_i}(x),$$

where $p_{X|A_i}(x) = \mathbb{P}(X = x|A_i)$ is the conditional PMF for X at x given A_i. The total probability theorem for a continuous RV X states:

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

where $f_{X|A_i}(x) \approx \frac{1}{\delta}\mathbb{P}(X \in [x, x+\delta]|A_i)$ is the conditional PDF for X at x given A_i.

Conditioning an RV on an event

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 $X \sim \mathrm{Uni}[10, 30]$ be the time of our arrival at the train station.
- Let Y be the waiting time to board the train; note $Y \in [0, 15]$.
- Define $A = \{X \in [10, 15]\}$ and $A^c = \{X \in [15, 30]\}$.
- Conditioned on A, Y is uniform over 0 and 5 while conditioned on A^c, Y is uniform over $[0, 15]$. Thus:

$$
\begin{aligned}
f_Y(y) &= \mathbb{P}(A) f_{Y|A}(y) + \mathbb{P}(A^c) f_{Y|A^c}(y) \\
&= \begin{cases} \frac{1}{4}\frac{1}{5} + \frac{3}{4}\frac{1}{15} = \frac{1}{10}, & y \in [0,5] \\ \frac{1}{4} 0 + \frac{3}{4}\frac{1}{15} = \frac{1}{20}, & y \in (5,15] \end{cases}
\end{aligned}
$$

Outline

Conditioning one RV on another

Recall the definition of the conditional PMF $p_{X|Y}(x|y)$ is the ratio of the joint over the marginal:

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}.$$

Let (X, Y) be a pair of continuous RVs with joint PDF $f_{X,Y}(x,y)$. Then the conditional PDF of X given $Y = y$ is:

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

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:

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

Conditioning one RV on another

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

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi r^2}, & x^2 + y^2 \leq r^2 \\ 0, & \text{else} \end{cases}$$

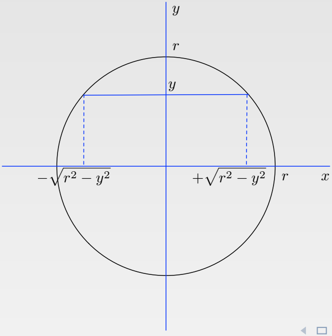
Find the conditional distribution $f_{X|Y}(x|y)$.
First find the marginal PDF $f_Y(y)$:

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

Conditioning one RV on another

The conditional PDF is:

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



Conditioning one RV on another

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

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

Thus:

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

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

Conditioning one RV on another

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)?

Conditioning one RV on another

First:

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

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

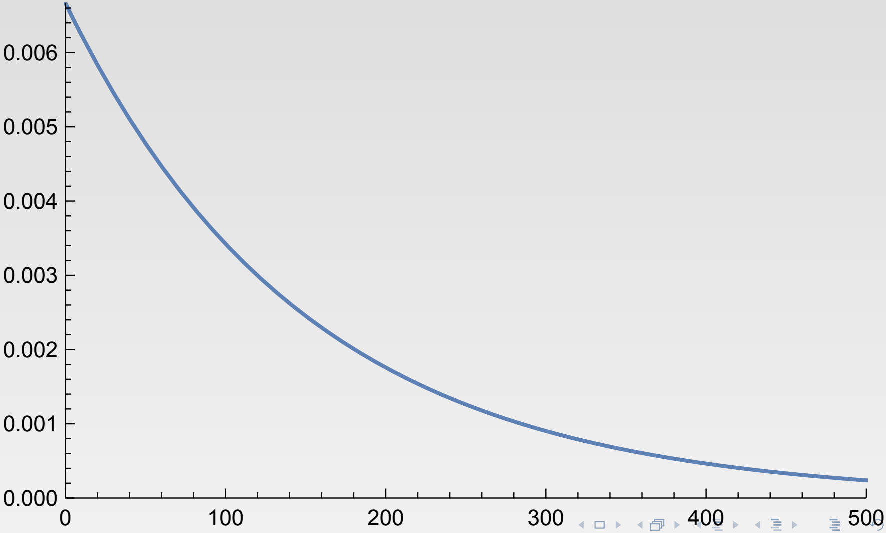
$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}(x/10)} \mathrm{e}^{-\frac{(y-x)^2}{2x^2/100}}.$$

Then:

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

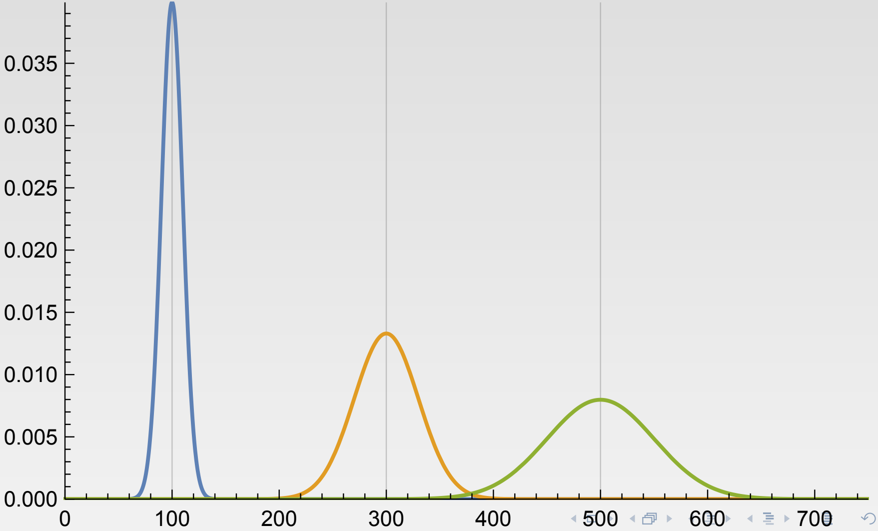
Conditioning one RV on another

Example. The marginal distribution f_X.



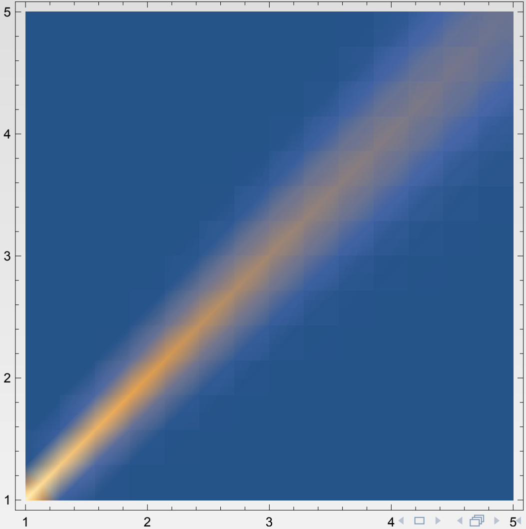
Conditioning one RV on another

Example. The conditional distribution $f_{Y|X}(y|x)$ for $x \in \{100, 300, 500\}$.



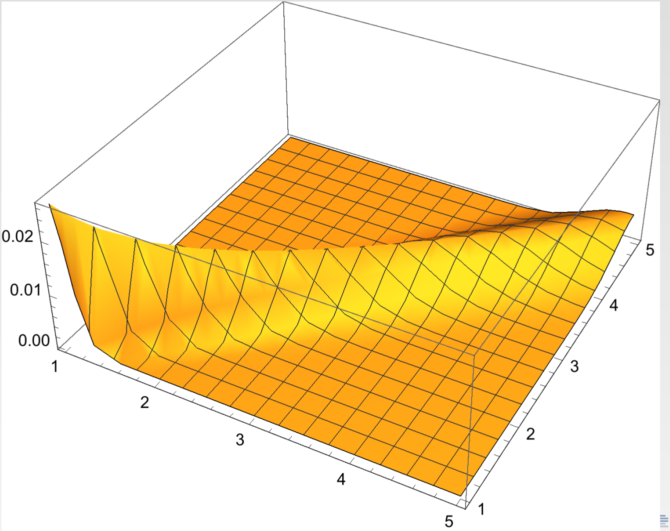
Conditioning one RV on another

Example. The joint distribution $f_{X,Y}(x, y)$ (contour plot).



Conditioning one RV on another

Example. The joint distribution $f_{X,Y}(x, y)$ (3D plot).



Outline

Conditional expectation

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

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

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

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

For a function $g(X)$:

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

Conditional expectation

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

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

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

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

Analogously for $g(X)$:

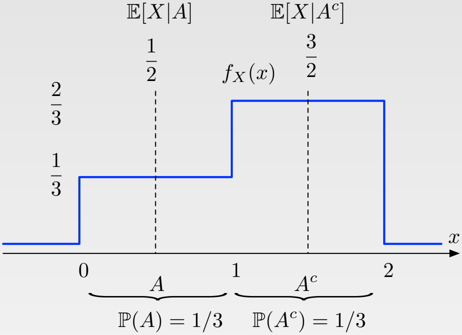
$$\begin{aligned} \mathbb{E}[g(X, Y)|Y = y] &= \int g(x, y) f_{X|Y}(x|y) \mathrm{d}x \\ \mathbb{E}[g(X, Y)] &= \int \mathbb{E}[g(X, Y)|Y = y] f_Y(y) \mathrm{d}y. \end{aligned}$$

Conditional expectation

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

$$f_X(x) = \begin{cases} 1/3, & x \in [0,1] \\ 2/3, & x \in (1,2] \end{cases}$$

Find the mean and variance of X.



Conditional expectation

Example. (Mean and variance of a piecewise constant PDF.) Consider events $A = \{X \in [0,1]\}$ and $A^c = \{X \in (1,2]\}$. Then:

$$\mathbb{P}(A) = 1/3, \ \mathbb{P}(A^c) = 2/3.$$

The conditional distribution of X is:

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{\mathbb{P}(A)} = 1, & x \in [0,1] \\ 0, & x \in (1,2] \end{cases}$$

$$f_{X|A^c}(x) = \begin{cases} 0, & x \in [0,1] \\ \frac{f_X(x)}{\mathbb{P}(A)} = 1, & x \in (1,2] \end{cases}$$

Further:

$$\mathbb{E}[X|A] = \int_0^1 x f_{X|A}(x)\mathrm{d}x = \frac{1}{2}$$

$$\mathbb{E}[X|A^c] = \int_1^2 x f_{X|A^c}(x)\mathrm{d}x = \frac{3}{2}$$

Conditional expectation

Example. (Mean and variance of a piecewise constant PDF.) This allows:

$$\mathbb{E}[X] = \mathbb{E}[X|A]\mathbb{P}(A) + \mathbb{E}[X|A^c]\mathbb{P}(A^c) = \frac{1}{2} \times \frac{1}{3} + \frac{3}{2} \times \frac{2}{3} = \frac{7}{6}.$$

To find the variance we first find $\mathbb{E}[X^2]$, and to do that we again condition on A:

$$\begin{aligned} \mathbb{E}[X^2|A] &= \int_0^1 x^2 f_{X|A}(x)\mathrm{d}x = \int_0^1 x^2 f_{X|A}(x)\mathrm{d}x = \frac{1}{3} x^3\Big|_0^1 = \frac{1}{3} \\ \mathbb{E}[X^2|A^c] &= \int_1^2 x^2 f_{X|A^c}(x)\mathrm{d}x = \int_1^2 x^2 f_{X|A^c}(x)\mathrm{d}x = \frac{1}{3} x^3\Big|_1^2 = \frac{7}{3} \end{aligned}$$

Then:

$$\mathbb{E}[X^2] = \mathbb{E}[X^2|A]\mathbb{P}(A) + \mathbb{E}[X^2|A^c]\mathbb{P}(A^c) = \frac{1}{3} \times \frac{1}{3} + \frac{7}{3} \times \frac{2}{3} = \frac{15}{9}$$

Thus the variance is

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

Outline

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 $f_Y(y) > 0$.
- This is directly analogous to the definition of independence for discrete RVs. There we said that (X,Y) are independent RVs if $p_{X|Y}(x|y) = p_X(x)$ for all (x,y) with $p_Y(y) > 0$.
- This in turn is derived from the definition of independent events (A,B). There we said that (A,B) are independent events if $\mathbb{P}(A|B) = \mathbb{P}(A)$, provided $\mathbb{P}(B) > 0$.

Independence

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

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

Independence

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

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

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

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

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

$$\begin{aligned}
\mathbb{E}[g(X)h(Y)] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x)h(y)f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x)h(y)f_X(x)f_Y(y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} g(x)f_X(x)\mathrm{d}x \int_{-\infty}^{+\infty} h(y)f_Y(y)\mathrm{d}y \\
&= \mathbb{E}[g(X)]\mathbb{E}[h(Y)].
\end{aligned}$$

Outline

The continuous Bayes' rule

Recall Bayes' rule for events.

- Event B is the effect (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:

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

The continuous Bayes' rule

Now we address Bayes' rule for continuous RVs.

- RV X is an unobserved phenomenon characterized by PDF $f_X(x)$
- RV Y is an observation with PDF f_Y related to X by a conditional PDF $f_{Y|X}(y|x)$
- 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:

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

The continuous Bayes' rule

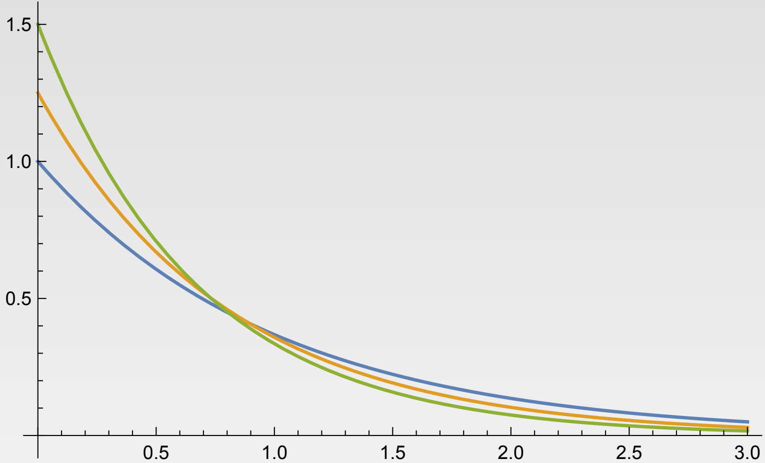
Example.

- A lightbulb lifetime Y is exponentially distributed with parameter λ.
- Due to quality control problems at the factory, λ is itself an RV, denoted $\Lambda \sim \mathrm{Uni}[1, 3/2]$.
- We test a lightbulb and record its lifetime – what can we say about the underlying parameter λ?
- The parameter distribution: $f_\Lambda(\lambda) = 2$, for $1 \leq \lambda \leq 3/2$
- The conditional parameter distribution:

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

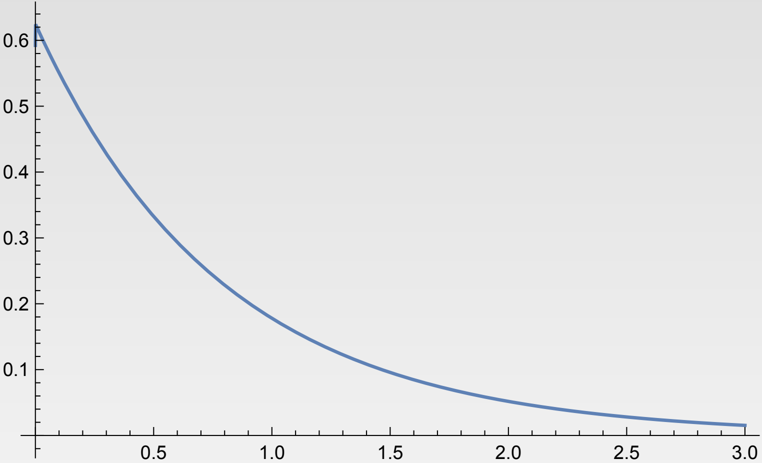
The continuous Bayes' rule

Example. Lightbulb lifetime. The conditional distribution $f_{Y|\Lambda}(y|\lambda)$ for $\lambda \in \{1, 5/4, 3/2\}$.



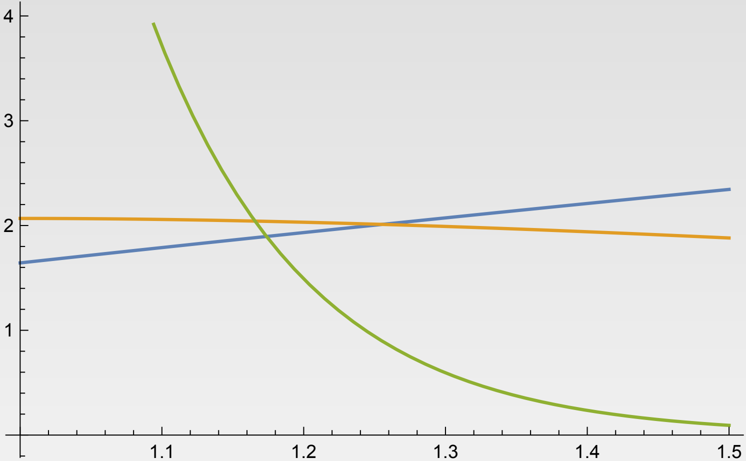
The continuous Bayes' rule

Example. Lightbulb lifetime. The unconditioned (marginal) distribution $f_Y(y) = \int_{\lambda_{\min}}^{\lambda_{\max}} f_{Y|\Lambda}(y|\lambda) f_\Lambda(\lambda) \mathrm{d}\lambda$.



The continuous Bayes' rule

Example. Lightbulb lifetime. The posterior distribution $f_{\Lambda|Y}(\lambda|y) = f_{Y|\Lambda}(y|\lambda)f_{\Lambda}(\lambda)/f_Y(y)$ for $y \in \{1/10, 1, 10\}$.



Outline

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:

$$\begin{aligned}
\mathbb{P}(A|Y = y) &\approx \mathbb{P}(A|y \leq Y \leq y + \delta) \\
&= \frac{\mathbb{P}(A)\mathbb{P}(y \leq Y \leq y + \delta|A)}{\mathbb{P}(y \leq Y \leq y + \delta)} \\
&= \frac{\mathbb{P}(A) f_{Y|A}(y)\delta}{f_Y(y)\delta} \\
&= \frac{\mathbb{P}(A) f_{Y|A}(y)}{f_Y(y)}
\end{aligned}$$

Inference about a discrete RV

Now use the TPT on the denominator conditioning on A:

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

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

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

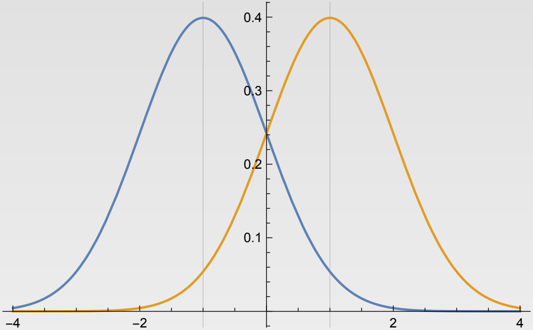
Inference about a discrete RV

Example. Signal detection.

- A binary signal $S \in \{-1, 1\}$ is transmitted with $\mathbb{P}(S = 1) = p = 1 - \mathbb{P}(S = -1)$ for some $p \in [0, 1]$.
- 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.
- Conditioned on $S = s$ we have $Y|\{S = s\} \sim N(s, 1)$.
- Find the probability that $S = +1$ given the observation y.

Inference about a discrete RV

Example. Signal detection. The two conditional observation distributions $f_{Y|S}(y|-1)$ (left) and $f_{Y|S}(y|+1)$ (right).



The obvious detection rule is guess $S = -1$ if $Y < 0$ or $S = +1$ if $Y > 0$. But this doesn't incorporate knowledge of p, the transmission probabilities.

Inference about a discrete RV

Example. Signal detection. By Bayes' rule:

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

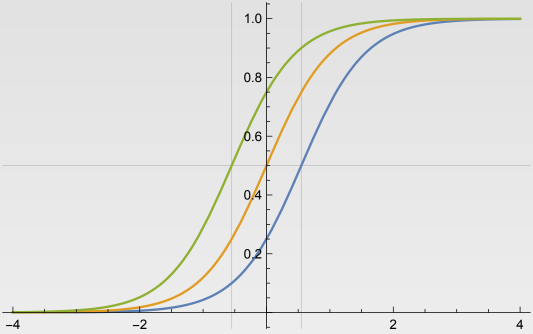
Solving $\mathbb{P}(S=+1|Y=y)=1/2$ for y gives

$$y(p)=\frac{1}{2}\log\left(\frac{1}{p}-1\right).$$

The new rule is: guess $S=+1$ if $y>y(p)$ or $S=-1$ if $y<y(p)$.

Inference about a discrete RV

Example. Signal detection. The probability of $S = +1$ given $Y = y$, $p_{S|Y}(+1|y)$, for $p \in \{1/4, 1/2, 3/4\}$.



The vertical gridlines give the value $y(p)$ such that $p_{S|Y}(+1|y) = 1/2$.

Inference about a discrete RV

Example. Signal detection. The threshold function $y(p)$ vs. $p \in [0, 1]$.