Bayes Rule for Discrete and Continuous Random Variables

Probability and Statistics for Data Science

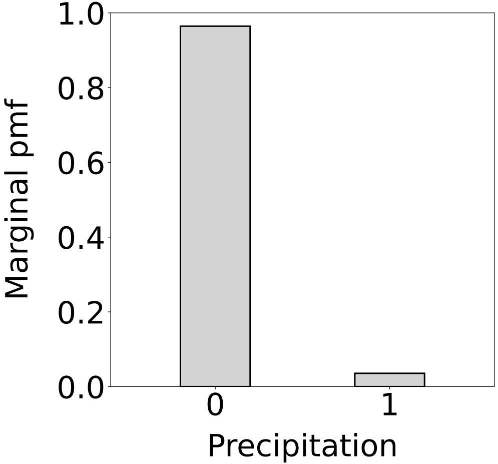
Carlos Fernandez-Granda

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

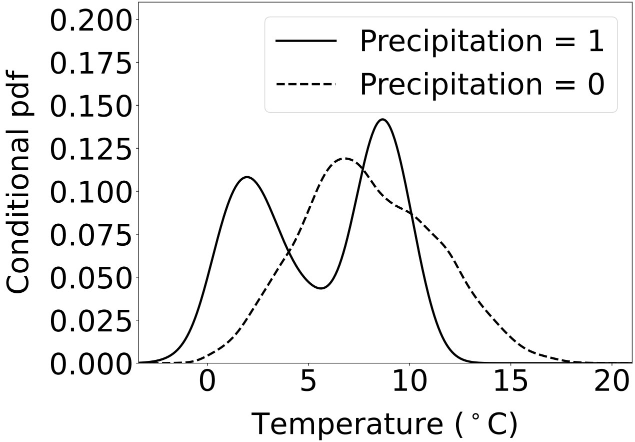
Mauna Loa

Temperature ($\tilde{c}$) and precipitation ($\tilde{d}$)

Marginal pmf of precipitation



Conditional pdf of temperature given precipitation



Conditional pmf

Conditional pmf of $\tilde{d}$ given $\tilde{c} = c$?

Problem: $\mathrm{P}(\tilde{c} = c) = 0$

As usual, we resort to limits

$$p_{\tilde{d} \mid \tilde{c}}(d \mid c) := \lim_{\epsilon \to 0} \mathrm{P}\left(\tilde{d} = d \mid c - \epsilon < \tilde{c} \leq c\right)$$

Marginal distribution of $\tilde{d}$

We know $f_{\tilde{c}}$ and $p_{\tilde{d}\,|\,\tilde{c}}(\cdot\,|\,c)$ for all c

Marginal distribution of $\tilde{d}$?

$$p_{\tilde{d}}(d) = \int_{c=-\infty}^{\infty} f_{\tilde{c}}(c)\, p_{\tilde{d}\,|\,\tilde{c}}(d\,|\,c)\, \mathrm{d}c$$

Sketch of proof

Grid $\{..., c_{-1}, c_0, c_1, ...\}$ with step size ϵ

$$p_{\tilde{d}}(d) = \sum_{i=-\infty}^{\infty} \mathrm{P}\left(\tilde{d} = d, c_i - \epsilon < \tilde{c} \leq c_i\right)$$

As $\epsilon \to 0$

$$\begin{aligned}
p_{\tilde{d}}(d) &= \int_{c=-\infty}^{\infty} \lim_{\epsilon \to 0} \frac{\mathrm{P}\left(\tilde{d} = d, c - \epsilon < \tilde{c} \leq c\right)}{\epsilon} \mathrm{d}c \\
&= \int_{c=-\infty}^{\infty} \lim_{\epsilon \to 0} \frac{\mathrm{P}\left(c - \epsilon < \tilde{c} \leq c\right)}{\epsilon} \cdot \frac{\mathrm{P}\left(\tilde{d} = d, c - \epsilon < \tilde{c} \leq c\right)}{\mathrm{P}\left(c - \epsilon < \tilde{c} \leq c\right)} \mathrm{d}c \\
&= \int_{c=-\infty}^{\infty} f_{\tilde{c}}(c)\, p_{\tilde{d} \mid \tilde{c}}(d \mid c)\, \mathrm{d}c
\end{aligned}$$

Chain rule

For discrete $\tilde{a}$ and $\tilde{b}$

$$\begin{aligned} p_{\tilde{a},\tilde{b}}(a,b) &= p_{\tilde{a}}(a)\, p_{\tilde{b}\,|\,\tilde{a}}(b\,|\,a) \\ &= p_{\tilde{b}}(b)\, p_{\tilde{a}\,|\,\tilde{b}}(a\,|\,b) \end{aligned}$$

For continuous $\tilde{a}$ and $\tilde{b}$

$$\begin{aligned} f_{\tilde{a},\tilde{b}}(a,b) &= f_{\tilde{a}}(a)\, f_{\tilde{b}\,|\,\tilde{a}}(b\,|\,a) \\ &= f_{\tilde{b}}(b)\, f_{\tilde{a}\,|\,\tilde{b}}(a\,|\,b) \end{aligned}$$

For discrete $\tilde{d}$ and continuous $\tilde{c}$?

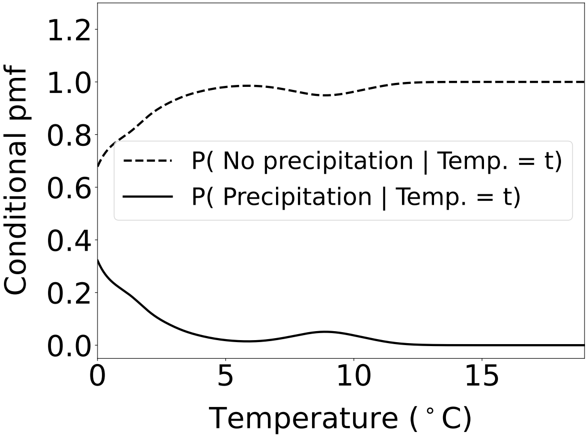
$$p_{\tilde{d}}(d)\, f_{\tilde{c}\,|\,\tilde{d}}(c\,|\,d) = f_{\tilde{c}}(c)\, p_{\tilde{d}\,|\,\tilde{c}}(d\,|\,c) \quad ?$$

Chain rule

$$
\begin{aligned}
p_{\tilde{d}}(d)\, f_{\tilde{c}\,|\,\tilde{d}}(c\,|\,d) &= \mathrm{P}(\tilde{d}=d) \lim_{\epsilon\to 0} \frac{\mathrm{P}\left(c-\epsilon < \tilde{c} \le c \,|\, \tilde{d}=d\right)}{\epsilon} \\
&= \lim_{\epsilon\to 0} \frac{\mathrm{P}(\tilde{d}=d)\mathrm{P}\left(c-\epsilon < \tilde{c} \le c \,|\, \tilde{d}=d\right)}{\epsilon} \\
&= \lim_{\epsilon\to 0} \frac{\mathrm{P}\left(\tilde{d}=d, c-\epsilon < \tilde{c} \le c\right)}{\epsilon} \\
&= \lim_{\epsilon\to 0} \frac{\mathrm{P}\left(c-\epsilon < \tilde{c} \le c\right)}{\epsilon} \mathrm{P}\left(\tilde{d}=d \,|\, c-\epsilon < \tilde{c} \le c\right) \\
&= f_{\tilde{c}}(c)\, p_{\tilde{d}\,|\,\tilde{c}}(d\,|\,c)
\end{aligned}
$$

Mauna Loa

$$p_{\tilde{d}\,|\,\tilde{c}}\,(d\,|\,c) = \frac{p_{\tilde{d}}\,(d)\,f_{\tilde{c}\,|\,\tilde{d}}\,(c\,|\,d)}{f_{\tilde{c}}(c)}$$



Gaussian mixture model

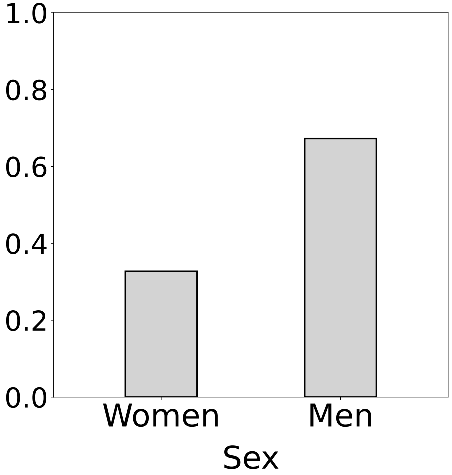
Height: Continuous random variable $\tilde{h}$

Sex: Discrete random variable $\tilde{s}$

Conditional distribution of $\tilde{h}$ given $\tilde{s}$ is Gaussian

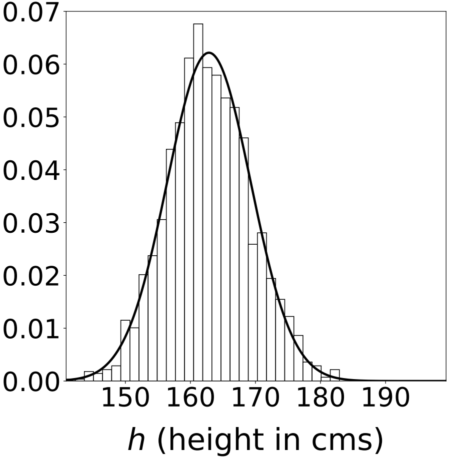
Marginal distribution of $\tilde{s}$

1,986 women and 4,082 men



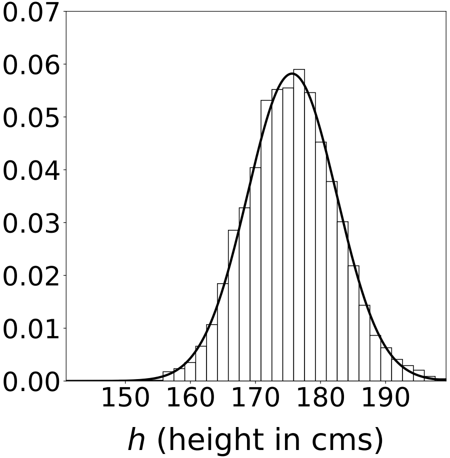
Conditional distribution of $\tilde{h}$ given $\tilde{s} = \text{woman}$

Gaussian with $\mu_{\text{women}} = 163$ cm and $\sigma_{\text{women}} = 6.4$ cm



Conditional distribution of $\tilde{h}$ given $\tilde{s}$ = man

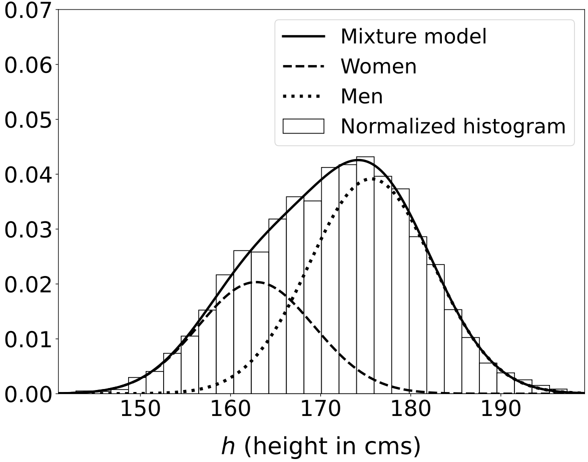
Gaussian with $\mu_{\text{men}} = 176$ cm and $\sigma_{\text{men}} = 6.9$ cm



Marginal distribution of $\tilde{h}$

$$
\begin{aligned}
f_{\tilde{h}}(h) &= \sum_{s=0}^{1} p_{\tilde{s}}(s)\, f_{\tilde{h}\,|\,\tilde{s}}(h\,|\,s) \\
&= \frac{\pi_{\text{women}}}{\sqrt{2\pi}\sigma_{\text{women}}} \exp\left(-\frac{1}{2}\left(\frac{h-\mu_{\text{women}}}{\sigma_{\text{women}}}\right)^2\right) \\
&\quad + \frac{\pi_{\text{men}}}{\sqrt{2\pi}\sigma_{\text{men}}} \exp\left(-\frac{1}{2}\left(\frac{h-\mu_{\text{men}}}{\sigma_{\text{men}}}\right)^2\right)
\end{aligned}
$$

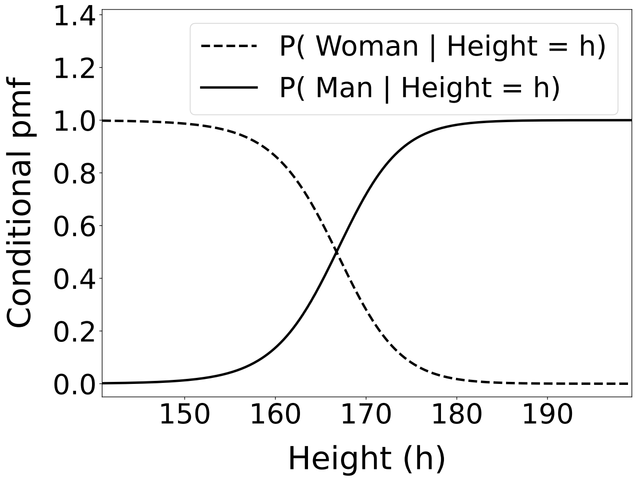
Gaussian mixture model



Conditional distribution of $\tilde{s}$ given $\tilde{h}$?

$$
\begin{aligned}
& p_{\tilde{s} \mid \tilde{h}}(0 \mid h) \\
&= \frac{p_{\tilde{s}}(0) f_{\tilde{h} \mid \tilde{s}}(h \mid 0)}{p_{\tilde{s}}(0) f_{\tilde{h} \mid \tilde{s}}(h \mid 0) + p_{\tilde{s}}(1) f_{\tilde{h} \mid \tilde{s}}(h \mid 1)} \\
&= \frac{\frac{\theta}{\sqrt{2\pi}\sigma_{\textbf{women}}} \exp\left(-\frac{1}{2}\left(\frac{h-\mu_{\textbf{women}}}{\sigma_{\textbf{women}}}\right)^2\right)}{\frac{\theta}{\sqrt{2\pi}\sigma_{\textbf{women}}} \exp\left(-\frac{1}{2}\left(\frac{h-\mu_{\textbf{women}}}{\sigma_{\textbf{women}}}\right)^2\right) + \frac{1-\theta}{\sqrt{2\pi}\sigma_{\textbf{men}}} \exp\left(-\frac{1}{2}\left(\frac{h-\mu_{\textbf{men}}}{\sigma_{\textbf{men}}}\right)^2\right)} \\
&= \frac{1}{1 + \frac{1-\theta}{\theta} \frac{\sigma_{\textbf{women}}}{\sigma_{\textbf{men}}} \exp\left(\frac{1}{2}\left(\frac{h-\mu_{\textbf{women}}}{\sigma_{\textbf{women}}}\right)^2 - \frac{1}{2}\left(\frac{h-\mu_{\textbf{men}}}{\sigma_{\textbf{men}}}\right)^2\right)} \\
&= \frac{1}{1 + 0.7 \exp\left(0.0017 h^2 - 0.28 h\right)}
\end{aligned}
$$

Conditional pmf of $\tilde{s}$ given $\tilde{h}$

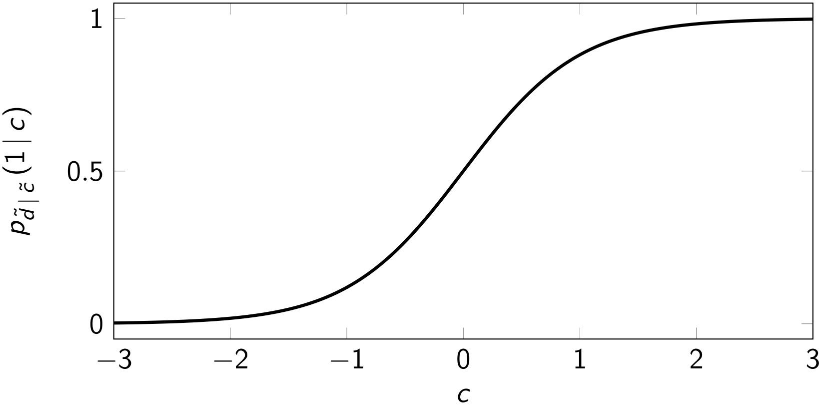


Gaussian mixture model with fixed variance

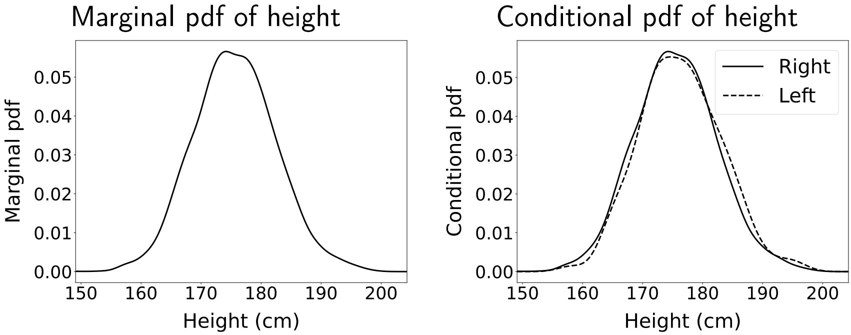
Means μ_1 and μ_2 are different, but variance σ^2 is the same

$$
\begin{aligned}
p_{\tilde{s}\,|\,\tilde{c}}(1\,|\,c) &= \frac{p_{\tilde{s}}(1)\, f_{\tilde{c}\,|\,\tilde{s}}(c\,|\,1)}{p_{\tilde{s}}(0)\, f_{\tilde{c}\,|\,\tilde{s}}(c\,|\,0) + p_{\tilde{s}}(1)\, f_{\tilde{c}\,|\,\tilde{s}}(c\,|\,1)} \\
&= \frac{\frac{\theta}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{c-\mu_1}{\sigma}\right)^2\right)}{\frac{\theta}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{c-\mu_1}{\sigma}\right)^2\right) + \frac{1-\theta}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{c-\mu_0}{\sigma}\right)^2\right)} \\
&= \frac{1}{1 + \frac{1-\theta}{\theta} \exp\left(\frac{1}{2}\left(\frac{c-\mu_1}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{c-\mu_0}{\sigma}\right)^2\right)} \\
&= \frac{1}{1 + \frac{1-\theta}{\theta} \exp\left(\frac{\mu_0-\mu_1}{\sigma^2} c + \frac{1}{2\sigma^2}\left(\mu_1^2 - \mu_0^2\right)\right)} \\
&= \frac{1}{1 + \alpha \exp(-\beta c)}
\end{aligned}
$$

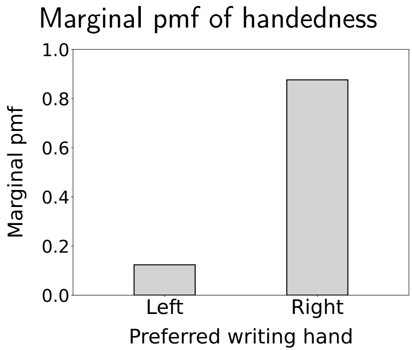
Logistic function

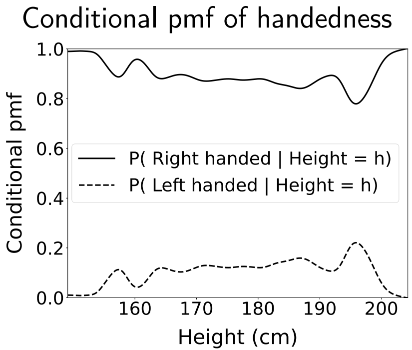


Height and handedness



Height and handedness





Independence

Two random variables $\tilde{c}$ and $\tilde{d}$ are independent if our uncertainty about $\tilde{c}$ does not change when information about $\tilde{d}$ is revealed

Independence

For a continuous random variable $\tilde{c}$ and a discrete random variable $\tilde{d}$, for any possible values of c and d we should have

$$p_{\tilde{d} \mid \tilde{c}}(d \mid c) = p_{\tilde{d}}(d)$$

$$f_{\tilde{c} \mid \tilde{d}}(c \mid d) = f_{\tilde{c}}(c)$$

Conditional independence

Two random variables $\tilde{c}$ and $\tilde{d}$ are conditionally independent given $\tilde{a}$ if our uncertainty about $\tilde{c}$ does *not* change when information about $\tilde{d}$ is revealed, as long as the value of $\tilde{a}$ is known

Conditional independence

A pair of continuous and discrete random variables $\tilde{c}$ and $\tilde{d}$ are conditionally independent given $\tilde{a}$ if and only if

$$\begin{aligned} p_{\tilde{d} \mid \tilde{c}, \tilde{a}}(d \mid c, a) &= p_{\tilde{d} \mid \tilde{a}}(d \mid a) \\ f_{\tilde{c} \mid \tilde{d}, \tilde{a}}(c \mid d, a) &= f_{\tilde{c} \mid \tilde{a}}(c \mid a) \quad \text{for all } a, c, d \end{aligned}$$

What have we learned?

Bayes rule for discrete and continuous random variables

Definition of independence between discrete and continuous variables