The Conditional Mean And Iterated Expectation

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

Motivation

We are interested in the mean of $\tilde{b}$

We have access to the conditional mean function of $\tilde{b}$ given $\tilde{a}$

Conditional mean function

The conditional mean function of a discrete random variable $\tilde{b}$ given $\tilde{a}$ is

$$\mu_{\tilde{b} \,|\, \tilde{a}}(a) := \sum_{b \in B} b \, p_{\tilde{b} \,|\, \tilde{a}}(b \,|\, a)$$

How can we compute $\mathrm{E}[\tilde{b}]$ from $\mu_{\tilde{b} \,|\, \tilde{a}}$?

Intuitive definition of mean

Data: y_1, y_2, ..., y_n

Interpreted as samples from $\tilde{b}$

$$\mathrm{E}[\tilde{b}] \approx \frac{1}{n} \sum_{i=1}^{n} y_i$$

Intuitive definition of conditional mean

Dataset $\mathcal{D}$: (x_1, y_1), (x_2, y_2), ..., (x_n, y_n), where $x_i \in A$

Data interpreted as samples from random variables $\tilde{a}$ (range A) and $\tilde{b}$

$$Y_a := \{y \mid (a, y) \in \mathcal{D}\}$$

$$\mu_{\tilde{b} \mid \tilde{a}}(a) \approx \frac{1}{n_a} \sum_{y \in Y_a} y$$

n_a = number of elements of Y_a

Intuitive definition of probability

$$\mathrm{P}(\tilde{a} = a) = \frac{n_a}{n}$$

Intuitive definition of mean

Data: $y_1, y_2, \ldots, y_n$

Interpreted as samples from $\tilde{b}$

$$\mathrm{E}[\tilde{b}] \approx \frac{1}{n} \sum_{i=1}^{n} y_i$$

Mean and conditional mean function

$$
\begin{aligned}
\mathrm{E}[\tilde{b}] &\approx \frac{1}{n} \sum_{i=1}^{n} y_i \\
&= \frac{1}{n} \sum_{a \in A} \sum_{y \in Y_a} y \\
&= \sum_{a \in A} \frac{n_A}{n} \frac{1}{n_A} \sum_{y \in Y_a} y \\
&\approx \sum_{a \in A} \frac{n_A}{n} \mu_{\tilde{b} \mid \tilde{a}}(a) \\
&\approx \sum_{a \in A} p_{\tilde{a}}(a) \mu_{\tilde{b} \mid \tilde{a}}(a) \\
&= \sum_{a \in A} p_{\tilde{a}}(a) h(a) \qquad \text{for } h := \mu_{\tilde{b} \mid \tilde{a}}
\end{aligned}
$$

Function of a random variable

The mean of $h(\tilde{a})$, $h : \mathbb{R} \to \mathbb{R}$ is

$$\mathrm{E}\left[h(\tilde{a})\right] := \sum_{a \in A} h(a)\, p_{\tilde{a}}(a)$$

if $\tilde{a}$ is discrete and the sum converges

Iterated expectation

$$
\begin{aligned}
\mathrm{E}[\tilde{b}] &\approx \sum_{a \in A} p_{\tilde{a}}(a) \mu_{\tilde{b} \mid \tilde{a}}(a) \\
&= \mathrm{E}\left[\mu_{\tilde{b} \mid \tilde{a}}(\tilde{a})\right]
\end{aligned}
$$

Mean and conditional mean function

If $\tilde{a}$ is continuous, a similar argument yields

$$\begin{aligned} \mathrm{E}[\tilde{b}] &\approx \int_{a=-\infty}^{\infty} f_{\tilde{a}}(a)\mu_{\tilde{b}\,|\,\tilde{a}}(a)\,\mathrm{d}a \\ &= \int_{a=-\infty}^{\infty} f_{\tilde{a}}(a)h(a)\,\mathrm{d}a \qquad \text{for } h := \mu_{\tilde{b}\,|\,\tilde{a}} \end{aligned}$$

Function of a random variable

The mean of $h(\tilde{a})$, $h : \mathbb{R} \to \mathbb{R}$ is

$$\mathrm{E}\left[h(\tilde{a})\right] := \int_{a=-\infty}^{\infty} h(a)\, f_{\tilde{a}}(a)\, \mathrm{d}a$$

if $\tilde{a}$ is continuous and the integral converges

Iterated expectation

$$
\begin{aligned}
\mathrm{E}[\tilde{b}] &\approx \int_{a=-\infty}^{\infty} f_{\tilde{a}}(a)\mu_{\tilde{b}\,|\,\tilde{a}}(a)\,\mathrm{d}a \\
&= \mathrm{E}\left[\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})\right]
\end{aligned}
$$

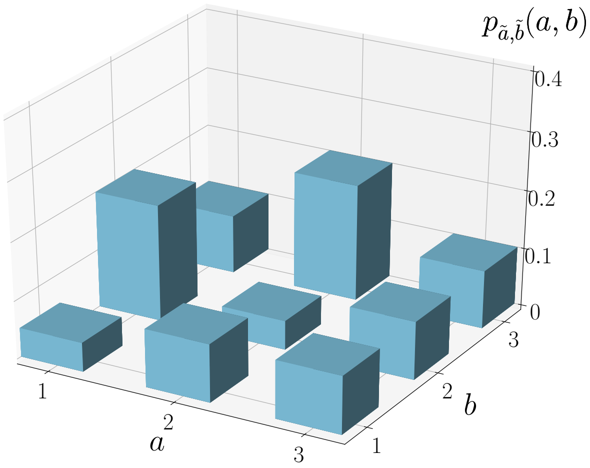
Conditional mean

The conditional mean function $\mu_{\tilde{b}\,|\,\tilde{a}}(a)$ is a deterministic function of a

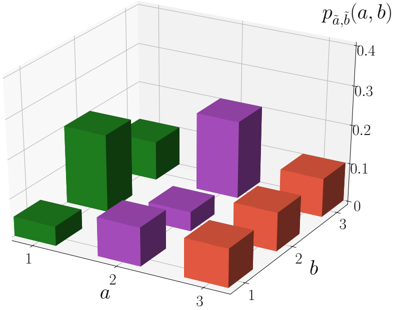
If we plug $\tilde{a}$ into it we obtain a random variable

We call $\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})$ the conditional mean

Joint pmf



Conditional mean function



$$\mu_{\tilde{b}\,|\,\tilde{a}}(1) = \sum_{b \in B} b \, p_{\tilde{b}\,|\,\tilde{a}}(b \,|\, 1) = 2.14$$

$$\mu_{\tilde{b}\,|\,\tilde{a}}(2) = \sum_{b \in B} b \, p_{\tilde{b}\,|\,\tilde{a}}(b \,|\, 2) = 2.57$$

$$\mu_{\tilde{b}\,|\,\tilde{a}}(3) = \sum_{b \in B} b \, p_{\tilde{b}\,|\,\tilde{a}}(b \,|\, a) = 2$$

Conditional mean

Distribution of $\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})$?

$$\mu_{\tilde{b}\,|\,\tilde{a}}(1) = \sum_{b \in B} b\, p_{\tilde{b}\,|\,\tilde{a}}(b\,|\,1) = \frac{15}{7}$$

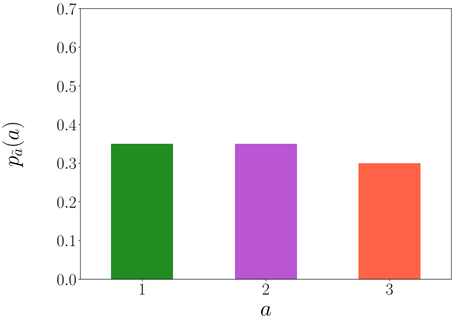
$$\mu_{\tilde{b}\,|\,\tilde{a}}(2) = \sum_{b \in B} b\, p_{\tilde{b}\,|\,\tilde{a}}(b\,|\,2) = \frac{16}{7}$$

$$\mu_{\tilde{b}\,|\,\tilde{a}}(3) = \sum_{b \in B} b\, p_{\tilde{b}\,|\,\tilde{a}}(b\,|\,a) = 2$$

We need the marginal pmf of $\tilde{a}$

Marginal pmf of the conditional mean

$$p_{\tilde{a}}(a) = \sum_{b=1}^{3} p_{\tilde{a},\tilde{b}}(a,b)$$



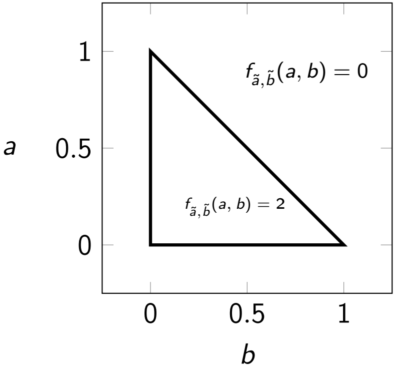
$$p_{\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})}\left(\frac{15}{7}\right) = \mathrm{P}\left(\tilde{a} = 1\right) = 0.35$$

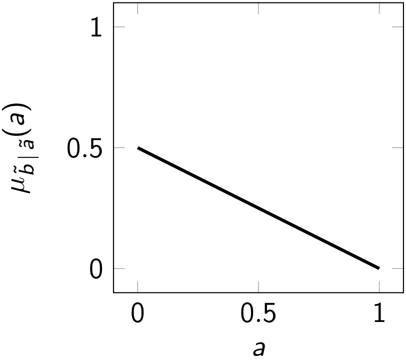
$$p_{\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})}\left(\frac{16}{7}\right) = \mathrm{P}\left(\tilde{a} = 2\right) = 0.35$$

$$p_{\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})}\left(2\right) = \mathrm{P}\left(\tilde{a} = 2\right) = 0.3$$

Triangle lake: Conditional mean function

$$\mu_{\tilde{b}\,|\,\tilde{a}}(a) = \int_{b=-\infty}^{\infty} b f_{\tilde{b}\,|\,\tilde{a}}(b\,|\,a)\,\mathrm{d}b$$
$$= \frac{1-a}{2}$$

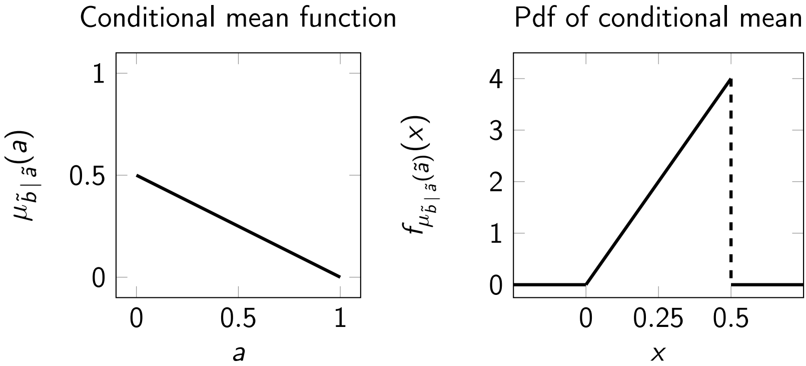




Triangle lake: Conditional mean

$$\begin{aligned}
\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a}) &= \frac{1-\tilde{a}}{2} \\
F_{\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})}(x) &= \mathrm{P}\left(\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a}) \leq x\right) \\
&= \mathrm{P}\left(\frac{1-\tilde{a}}{2} \leq x\right) \\
&= \mathrm{P}\left(\tilde{a} \geq 1-2x\right) \\
&= \int_{1-2x}^{1} f_{\tilde{a}}(t)\,\mathrm{d}t \\
&= \int_{1-2x}^{1} 2(1-t)\,\mathrm{d}t \\
&= 4x^2 \qquad \text{for } 0 \leq x \leq 0.5 \\
f_{\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})}(x) &= 8x
\end{aligned}$$

Triangle lake



What is the mean of the conditional mean?

$$
\begin{aligned}
\mathrm{E}[\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})] &= \int_{a=-\infty}^{\infty} f_{\tilde{a}}(a)\, \mu_{\tilde{b}\,|\,\tilde{a}}(a)\,\mathrm{d}a \\
&= \int_{a=-\infty}^{\infty} \int_{b=-\infty}^{\infty} f_{\tilde{a}}(a)\, f_{\tilde{b}\,|\,\tilde{a}}(b\,|\,a)\, b\,\mathrm{d}b\,\mathrm{d}a \\
&= \int_{a=-\infty}^{\infty} \int_{b=-\infty}^{\infty} f_{\tilde{a},\tilde{b}}(a,b)\, b\,\mathrm{d}b\,\mathrm{d}a \\
&= \mathrm{E}[\tilde{b}]
\end{aligned}
$$

Same for mean of function $h(\tilde{a}, \tilde{b})$

Same for discrete random variables

Iterated expectation

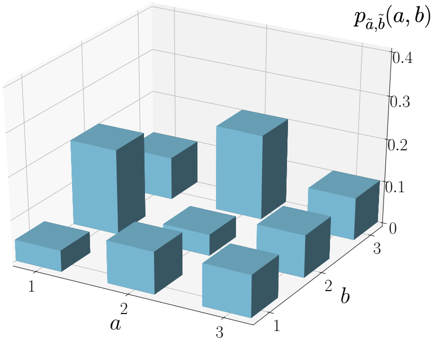
For any random variables $\tilde{a}$ and $\tilde{b}$ belonging to the same probability space

$$\mathrm{E}\left[\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})\right] = \mathrm{E}[\tilde{b}]$$

For any function $h : \mathbb{R}^2 \rightarrow \mathbb{R}$

$$\mathrm{E}[\mu_{h(\tilde{a},\tilde{b})\,|\,\tilde{a}}(\tilde{a})] = \mathrm{E}\left[h(\tilde{a},\tilde{b})\right]$$

Example



$$p_{\tilde{b}}(b) = \sum_{a=1}^{3} p_{\tilde{a},\tilde{b}}(a,b)$$

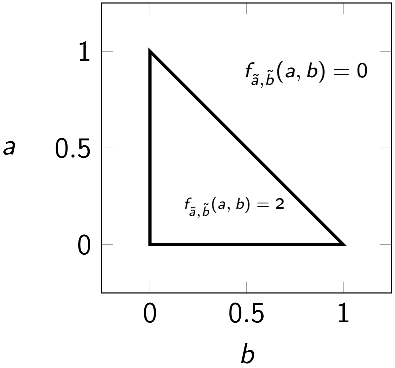
$$\mathrm{E}[\tilde{b}] = \sum_{b=1}^{3} b p_{\tilde{b}}(b) = 1 \cdot 0.25 + 2 \cdot 0.35 + 3 \cdot 0.4 = 2.15$$

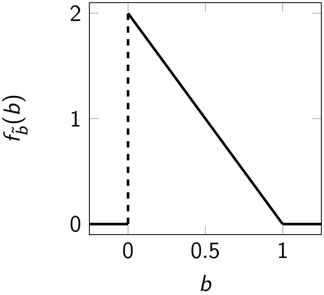
Iterated expectation

$$p_{\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})}\left(\frac{15}{7}\right) = 0.35 \qquad p_{\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})}\left(\frac{16}{7}\right) = 0.35 \qquad p_{\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})}(2) = 0.3$$

$$\begin{aligned} \mathrm{E}\left[\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})\right] &= \sum_{x \in \{2, 15/7, 16/7\}} x \, p_{\mu_{\tilde{b}\,|\,\tilde{a}}(\tilde{a})}(x) \\ &= 2 \cdot 0.3 + \frac{15}{7} \cdot 0.35 + \frac{16}{7} \cdot 0.35 \\ &= 2.15 = \mathrm{E}[\tilde{b}] \end{aligned}$$

Triangle lake: Conditional mean function



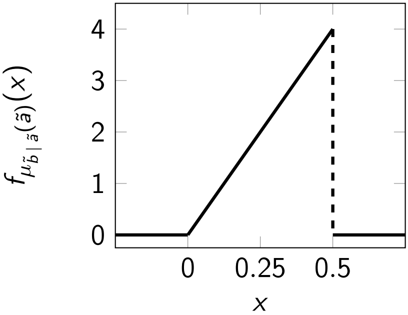


$$f_{\tilde{b}}(b) = \int_{a=-\infty}^{\infty} f_{\tilde{a},\tilde{b}}(a,b)\,\mathrm{d}a = 2(1-b)$$

$$\mathrm{E}[\tilde{b}] = \int_{b=-\infty}^{\infty} b f_{\tilde{b}}(b)\,\mathrm{d}b = \int_{b=0}^{1} 2b(1-b)\,\mathrm{d}b = \frac{1}{3}$$

Iterated expectation

Pdf of conditional mean



$$\mathrm{E}\left[\mu_{\tilde{b} \mid \tilde{a}}(\tilde{a})\right] = \int_{x=-\infty}^{\infty} x f_{\mu_{\tilde{b} \mid \tilde{a}}(\tilde{a})}(x) \, \mathrm{d}x$$

$$= \int_{x=0}^{\frac{1}{2}} 8x^2 \, \mathrm{d}x = \frac{1}{3} = \mathrm{E}[\tilde{b}]$$

Computer

Model for time $\tilde{t}$ until computer breaks down

Exponential random variable with parameter

$$\frac{1}{\tilde{o} + \tilde{c}}$$

$\tilde{o}$: fraction of time computer is off

$\tilde{c}$: how careful owner is

Both uniform in $[0, 1]$

Computer

Conditioned on $\tilde{o} = o$ and $\tilde{c} = c$, $\tilde{t}$ is exponential with parameter $\lambda := \frac{1}{o+c}$

$$\mu_{\tilde{t} \mid \tilde{o},\tilde{c}}(o, c) = \frac{1}{\lambda}$$
$$= o + c$$

$$\mathrm{E}\left[\tilde{t}\right] = \mathrm{E}\left[\mu_{\tilde{t} \mid \tilde{o},\tilde{c}}(\tilde{o}, \tilde{c})\right]$$
$$= \mathrm{E}\left[\tilde{o} + \tilde{c}\right]$$
$$= 0.5 + 0.5$$
$$= 1$$

Gaussian mixture model

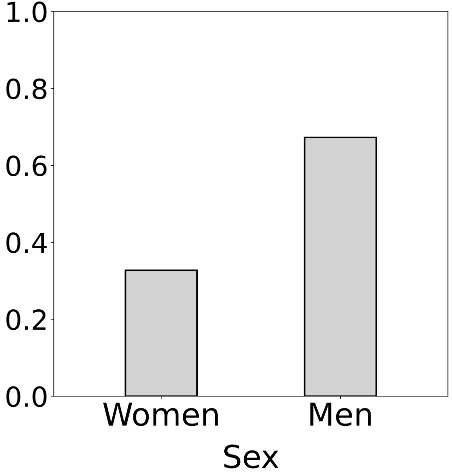
Height: Continuous random variable $\tilde{h}$

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

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

Marginal distribution of $\tilde{s}$

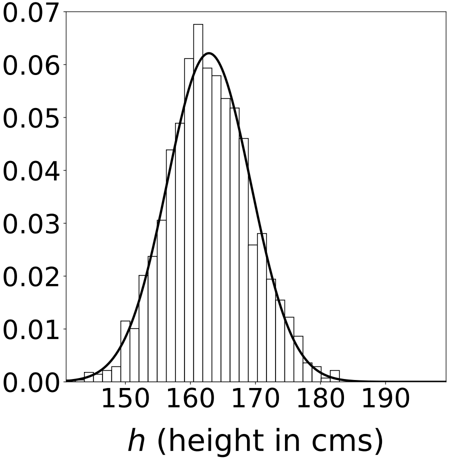
1,986 women and 4,082 men



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

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

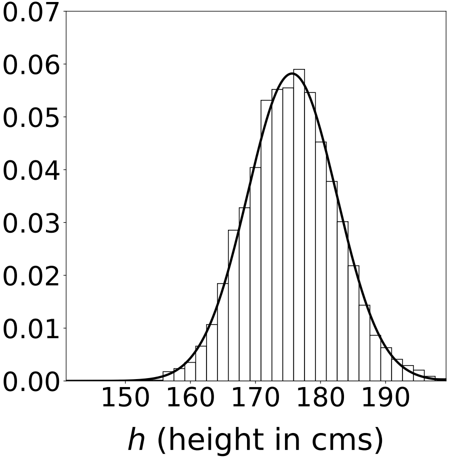
Conditional mean function $\mu_{\tilde{h}\,|\,\tilde{s}}(0) = 163$



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

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

Conditional mean function $\mu_{\tilde{h}\,|\,\tilde{s}}(1) = 176$



Iterated expection

$$
\begin{aligned}
\mathrm{E}[\tilde{h}] &= \mathrm{E}[\mu_{\tilde{h}\,|\,\tilde{s}}(\tilde{s})] \\
&= p_{\tilde{s}}(0)\mu_{\tilde{h}\,|\,\tilde{s}}(0) + p_{\tilde{s}}(1)\mu_{\tilde{h}\,|\,\tilde{s}}(1) \\
&= 171.7 \text{ cm}
\end{aligned}
$$

What have we learned

Definition of conditional mean as a random variable

Iterated expectation