4. Bayesian Inference, Part IV: Least Mean Squares (LMS) estimation

ECE 302 Fall 2009 TR 3-4:15pm

Purdue University, School of ECE

Prof. Ilya Pollak

An Estimate vs an Estimator

- Suppose X and Y are random variables.
- Y is observed, X is not observed.

An Estimate vs an Estimator

- Suppose X and Y are random variables.
- Y is observed, X is not observed.
- An estimator of X based on Y is any function $g(Y)$.

An Estimate vs an Estimator

- Suppose X and Y are random variables.
- Y is observed, X is not observed.
- An estimator of X based on Y is any function $g(Y)$.
- Since $g(Y)$ is a function of a r.v., it is a r.v. itself.

An Estimate vs an Estimator

- Suppose X and Y are random variables.
- Y is observed, X is not observed.
- An estimator of X based on Y is any function $g(Y)$.
- Since $g(Y)$ is a function of a r.v., it is a r.v. itself.
- An estimate of X based on $Y=y$ is a number, the value of an estimator as determined by the observed value y of Y.

An Estimate vs an Estimator

- Suppose X and Y are random variables.
- Y is observed, X is not observed.
- An estimator of X based on Y is any function $g(Y)$.
- Since $g(Y)$ is a function of a r.v., it is a r.v. itself.
- An estimate of X based on $Y=y$ is a number, the value of an estimator as determined by the observed value y of Y.
- E.g., if $g(Y)$ is an estimator of X, then $g(y)$ is an estimate of X for any specific observation y.

Pros and Cons of MAP Estimators

- Minimize the error probability.

Pros and Cons of MAP Estimators

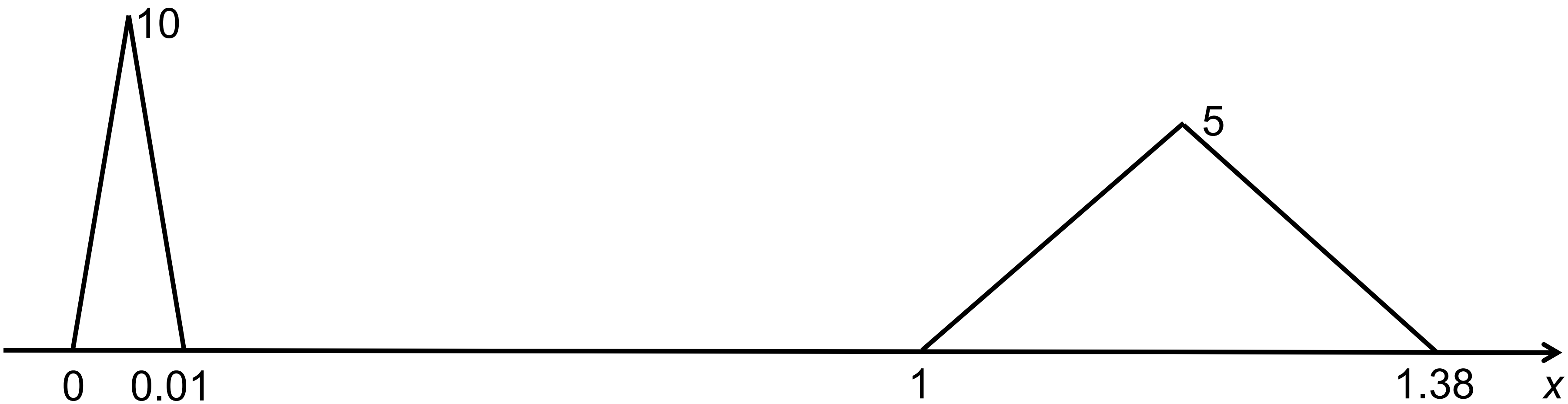
- Minimize the error probability.
- Not appropriate when different errors have different costs (as in, e.g., the burglar alarm and spam filtering examples).

Pros and Cons of MAP Estimators

- Minimize the error probability.
- Not appropriate when different errors have different costs (as in, e.g., the burglar alarm and spam filtering examples).
- For some posterior distributions, may result in large probabilities of large errors.

An example when the MAP estimate is bad

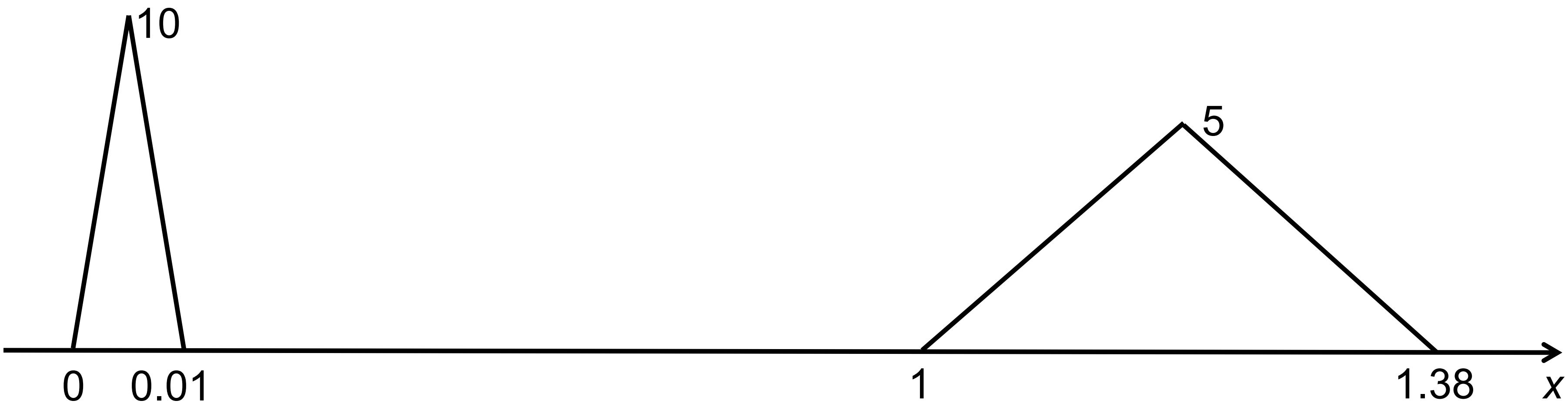
Suppose $f_{X|Y}(x|0)$ is:



- MAP: the max of the tall left triangle, $x^* = 0.005$.

An example when the MAP estimate is bad

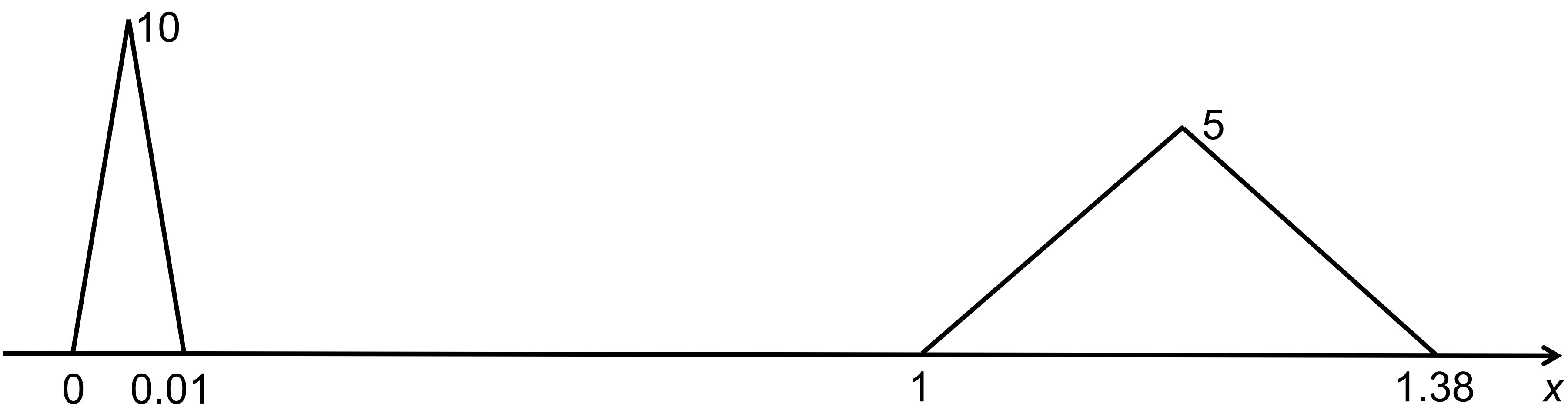
Suppose $f_{X|Y}(x|0)$ is:



- MAP: the max of the tall left triangle, $x^* = 0.005$.
- Left triangle: 0.05 of the conditional probability mass.

An example when the MAP estimate is bad

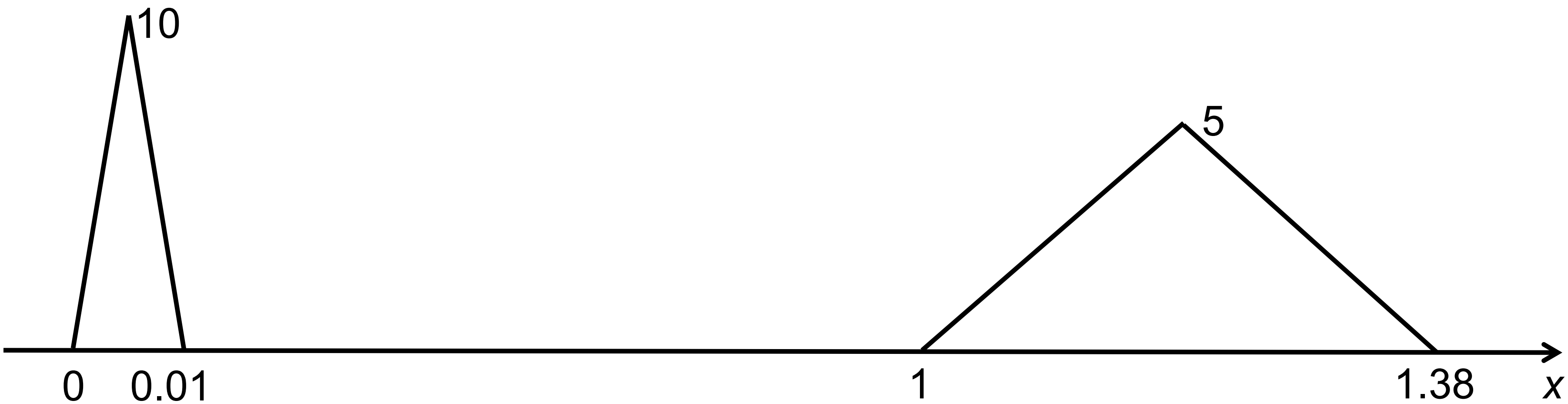
Suppose $f_{X|Y}(x|0)$ is:



- MAP: the max of the tall left triangle, $x^* = 0.005$.
- Left triangle: 0.05 of the conditional probability mass.
- MAP makes errors of ≥ 0.995 with cond. prob. 0.95.

An example when the MAP estimate is bad

Suppose $f_{X|Y}(x|0)$ is:



- MAP: the max of the tall left triangle, $x^* = 0.005$.
- Left triangle: 0.05 of the conditional probability mass.
- MAP makes errors of ≥ 0.995 with cond. prob. 0.95.
- Another estimate: conditional mean $E[X|Y=0] =$
  $0.05 \cdot 0.005 + 0.95 \cdot 1.19 \approx 1.13$.

Mean Squared Error (MSE) for an Estimator *g*(*Y*) of *X*

$$E\left[\left(X - g(Y)\right)^2\right]$$

Mean Squared Error (MSE) for an Estimator *g*(*Y*) of *X*

$$E\left[\left(X - g(Y)\right)^2\right]$$

- The expectation is taken with respect to the joint distribution of *X* and *Y*.

Mean Squared Error (MSE) for an Estimator *g*(*Y*) of *X*

$$E\left[\left(X - g(Y)\right)^2\right]$$

- The expectation is taken with respect to the joint distribution of *X* and *Y.*
- This is the expectation of the squared difference between the unobserved random variable *X* and its estimator based on the observed random variable *Y.*

Mean Squared Error (MSE) for an Estimator *g*(*Y*) of *X*

$$E\left[\left(X - g(Y)\right)^2\right]$$

- The expectation is taken with respect to the joint distribution of *X* and *Y.*
- This is the expectation of the squared difference between the unobserved random variable *X* and its estimator based on the observed random variable *Y.*
- Makes large errors more costly than small ones.

Conditional Mean Squared Error for an Estimate *g*(*y*) of *X*, given Y=*y*

$$E\left[\left(X - g(y)\right)^2 \middle| Y = y\right]$$

Conditional Mean Squared Error for an Estimate *g*(*y*) of *X*, given Y=*y*

$$E\left[\left(X - g(y)\right)^2 \middle| Y = y\right]$$

- The expectation is taken with respect to the conditional distribution of *X* given *Y=y.*

Conditional Mean Squared Error for an Estimate *g*(*y*) of *X*, given Y=*y*

$$E\left[\left(X-g(y)\right)^2 \middle| Y=y\right]$$

- The expectation is taken with respect to the conditional distribution of *X* given *Y=y.*
- This is the conditional expectation, given *Y=y*, of the squared difference between the unobserved random variable *X* and its estimate based on the specific observation *y* of the random variable *Y.*

Least Mean Squares (LMS) Estimator

- **LMS estimator for *X* based on *Y* is the estimator that minimizes the MSE among all estimators of *X* based on *Y*.**

Least Mean Squares (LMS) Estimator

- **LMS estimator for *X* based on *Y* is the estimator that minimizes the MSE among all estimators of *X* based on *Y*.**
- If $g(Y)$ is the LMS estimator, then, for any observation y of Y, the number $g(y)$ minimizes the conditional MSE among all estimates of X.

Least Mean Squares (LMS) Estimator

- **LMS estimator for X based on Y is the estimator that minimizes the MSE among all estimators of X based on Y.**
- If $g(Y)$ is the LMS estimator, then, for any observation y of Y, the number $g(y)$ minimizes the conditional MSE among all estimates of X.
- **The LMS estimator of X based on Y is $E[X|Y]$.**

Least Mean Squares (LMS) Estimator

- **LMS estimator for *X* based on *Y* is the estimator that minimizes the MSE among all estimators of *X* based on *Y*.**
- If $g(Y)$ is the LMS estimator, then, for any observation y of Y, the number $g(y)$ minimizes the conditional MSE among all estimates of X.
- **The LMS estimator of *X* based on Y is $E[X|Y]$.**
- For any observation y of Y, the LMS estimate of X is the mean of the posterior density, i.e., $E[X|Y=y]$.

LMS estimate with no data

- Suppose X is a r.v. with a known mean.

LMS estimate with no data

- Suppose X is a r.v. with a known mean.
- We observe *nothing.*

LMS estimate with no data

- Suppose X is a r.v. with a known mean.
- We observe *nothing.*
- We would like an estimate x^* of X which minimizes the MSE, $E[(X-x^*)^2]$.

LMS estimate with no data

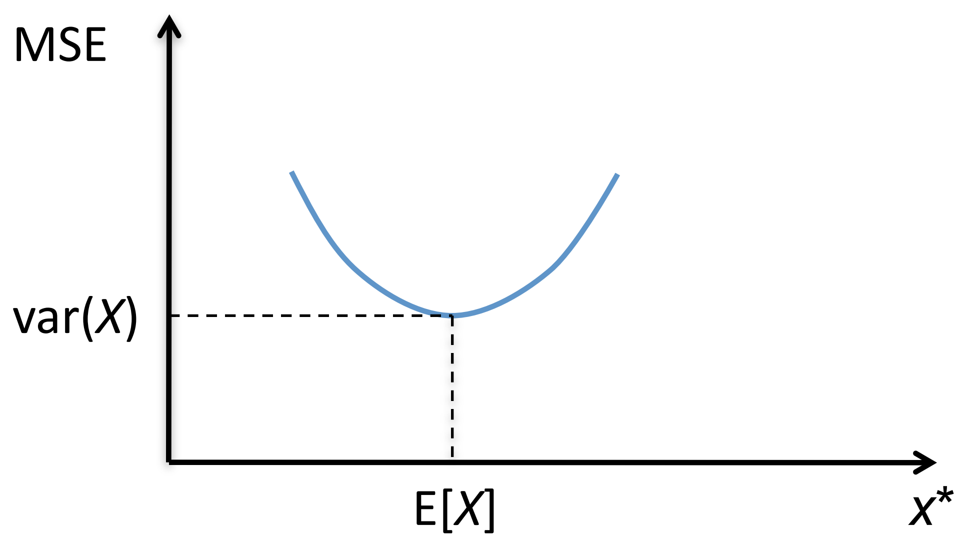
- Suppose X is a r.v. with a known mean.
- We observe *nothing.*
- We would like an estimate x^* of X which minimizes the MSE, $E[(X-x^*)^2]$.
- $E[(X-x^*)^2] = \text{var}(X-x^*) + (E[X-x^*])^2$

LMS estimate with no data

- Suppose X is a r.v. with a known mean.
- We observe *nothing.*
- We would like an estimate x^* of X which minimizes the MSE, $E[(X-x^*)^2]$.
- $E[(X-x^*)^2] = \mathrm{var}(X-x^*) + (E[X-x^*])^2 = \mathrm{var}(X) + (E[X] - x^*)^2$

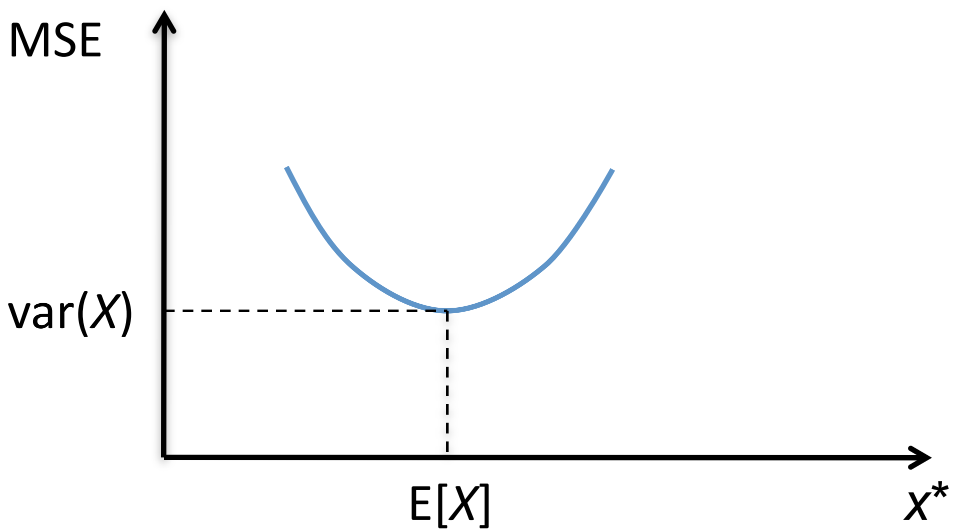
LMS estimate with no data

- Suppose X is a r.v. with a known mean.
- We observe *nothing.*
- We would like an estimate x^* of X which minimizes the MSE, $E[(X-x^*)^2]$.
- $E[(X-x^*)^2] = \mathrm{var}(X-x^*) + (E[X-x^*])^2 = \mathrm{var}(X) + (E[X] - x^*)^2$



LMS estimate with no data

- Suppose X is a r.v. with a known mean.
- We observe *nothing.*
- We would like an estimate x^* of X which minimizes the MSE, $E[(X-x^*)^2]$.
- $E[(X-x^*)^2] = \mathrm{var}(X-x^*) + (E[X-x^*])^2 = \mathrm{var}(X) + (E[X] - x^*)^2$
- This is a quadratic function of x^*, minimized by $x^*=E[X]$.



LMS estimate from one observation

- Suppose X, Y are r.v.'s, and $E[X|Y=y]$ is known.

LMS estimate from one observation

- Suppose X, Y are r.v.'s, and E[$X|Y=y$] is known.
- We observe $Y=y$.

LMS estimate from one observation

- Suppose X, Y are r.v.'s, and $E[X|Y=y]$ is known.
- We observe $Y=y$.
- We would like an estimate x^* of X which minimizes the posterior MSE, $E[(X-x^*)^2|Y=y]$.

LMS estimate from one observation

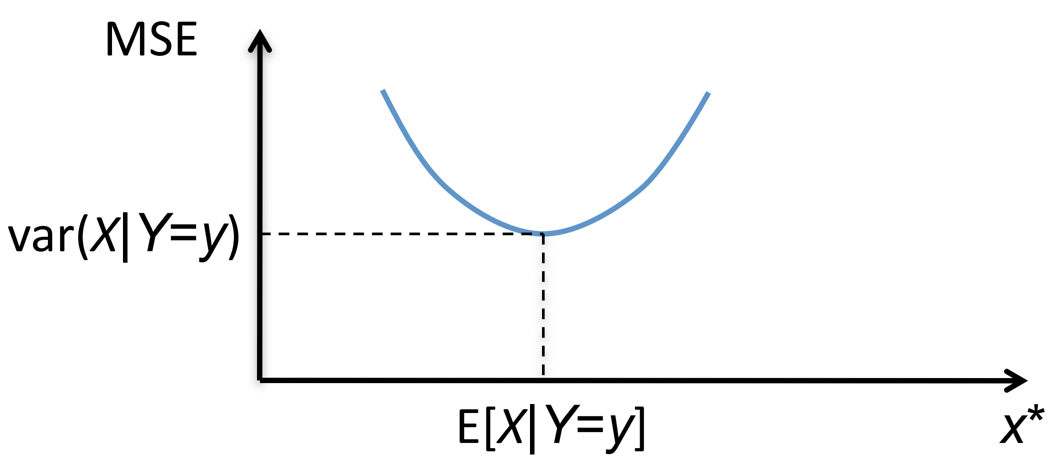
- Suppose X, Y are r.v.'s, and $E[X|Y=y]$ is known.
- We observe $Y=y$.
- We would like an estimate x^* of X which minimizes the posterior MSE, $E[(X-x^*)^2|Y=y]$.
- $E[(X-x^*)^2|Y=y] = \mathrm{var}(X-x^*|Y=y) + (E[X-x^*] \,|Y=y)^2$

LMS estimate from one observation

- Suppose X, Y are r.v.'s, and $E[X|Y=y]$ is known.
- We observe $Y=y$.
- We would like an estimate x^* of X which minimizes the posterior MSE, $E[(X-x^*)^2|Y=y]$.
- $E[(X-x^*)^2|Y=y] = \mathrm{var}(X-x^*|Y=y) + (E[X-x^*]\ |Y=y)^2$
  $= \mathrm{var}(X|Y=y) + (E[X|Y=y] - x^*)^2$

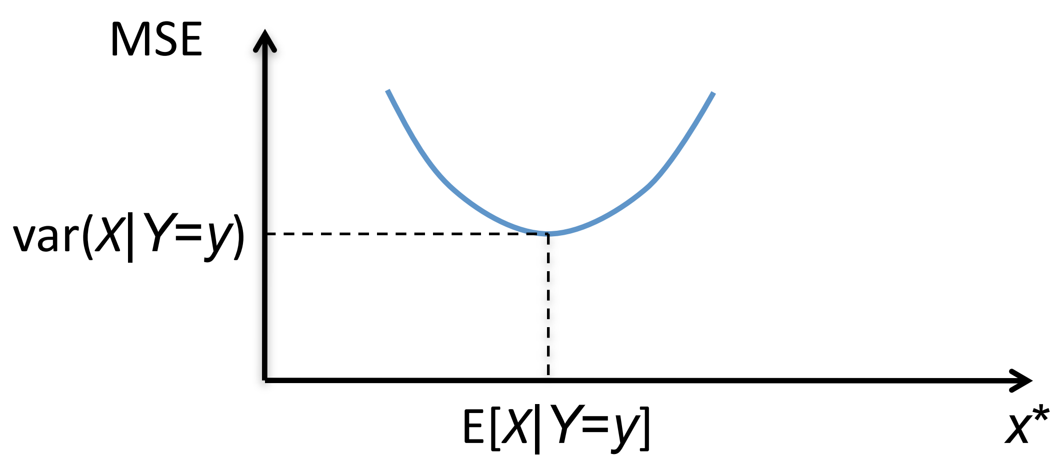
LMS estimate from one observation

- Suppose X, Y are r.v.'s, and $E[X|Y=y]$ is known.
- We observe $Y=y$.
- We would like an estimate x^* of X which minimizes the posterior MSE, $E[(X-x^*)^2|Y=y]$.
- $E[(X-x^*)^2|Y=y] = \mathrm{var}(X-x^*|Y=y) + (E[X-x^*]\ |Y=y)^2$
  $= \mathrm{var}(X|Y=y) + (E[X|Y=y] - x^*)^2$



LMS estimate from one observation

- Suppose X, Y are r.v.'s, and $E[X|Y=y]$ is known.
- We observe $Y=y$.
- We would like an estimate x^* of X which minimizes the posterior MSE, $E[(X-x^*)^2|Y=y]$.
- $E[(X-x^*)^2|Y=y] = \text{var}(X-x^*|Y=y) + (E[X-x^*]\ |Y=y)^2$
  $= \text{var}(X|Y=y) + (E[X|Y=y] - x^*)^2$
- Quadratic function of x^*, minimized by $x^*=E[X|Y=y]$.



LMS estimator

- Suppose X, Y are r.v.'s, and $E[X|Y=y]$ is known for all y.

LMS estimator

- Suppose X, Y are r.v.'s, and $E[X|Y=y]$ is known for all y.
- Let $g(y) = E[X|Y=y]$ be the LMS estimate of X based on data $Y=y$.

LMS estimator

- Suppose X, Y are r.v.'s, and E[X|Y=y] is known for all y.
- Let g(y) = E[X|Y=y] be the LMS estimate of X based on data Y=y.
- Then ***g*(*Y*) = E[*X*|*Y*]** is an estimator of X based on Y.

LMS estimator

- Suppose X, Y are r.v.'s, and $E[X|Y=y]$ is known for all y.
- Let $g(y) = E[X|Y=y]$ be the LMS estimate of X based on data $Y=y$.
- Then $\boldsymbol{g(Y) = E[X|Y]}$ is an estimator of X based on Y.
- Let $h(Y)$ be any other estimator of X based on Y.

LMS estimator

- Suppose X, Y are r.v.'s, and $E[X|Y=y]$ is known for all y.
- Let $g(y) = E[X|Y=y]$ be the LMS estimate of X based on data $Y=y$.
- Then $\boldsymbol{g(Y) = E[X|Y]}$ is an estimator of X based on Y.
- Let $h(Y)$ be any other estimator of X based on Y.
- We just showed that, for every value of y, we have:

  $E[(X-g(y))^2|Y=y] \le E[(X-h(y))^2|Y=y]$

LMS estimator

- Suppose X, Y are r.v.'s, and $E[X|Y=y]$ is known for all y.
- Let g(y) = $E[X|Y=y]$ be the LMS estimate of X based on data $Y=y$.
- Then $\boldsymbol{g(Y) = E[X|Y]}$ is an estimator of X based on Y.
- Let $h(Y)$ be any other estimator of X based on Y.
- We just showed that, for every value of y, we have:

  $E[(X-g(y))^2|Y=y] \le E[(X-h(y))^2|Y=y]$
- Therefore,

  $E[(X-g(Y))^2|Y] \le E[(X-h(Y))^2|Y]$

LMS estimator

- Suppose X, Y are r.v.'s, and $E[X|Y=y]$ is known for all y.
- Let g(y) = $E[X|Y=y]$ be the LMS estimate of X based on data $Y=y$.
- Then $\boldsymbol{g(Y) = E[X|Y]}$ is an estimator of X based on Y.
- Let $h(Y)$ be any other estimator of X based on Y.
- We just showed that, for every value of y, we have:

  $$E[(X-g(y))^2|Y=y] \le E[(X-h(y))^2|Y=y]$$
- Therefore,

  $$E[(X-g(Y))^2|Y] \le E[(X-h(Y))^2|Y]$$
- Take expectations of both sides and use the iterated expectation formula:

  $$E[(X-g(Y))^2] \le E[(X-h(Y))^2]$$

LMS estimator

- Suppose X, Y are r.v.'s, and $E[X|Y=y]$ is known for all y.
- Let g(y) = $E[X|Y=y]$ be the LMS estimate of X based on data $Y=y$.
- Then $\mathbf{g(Y) = E[X|Y]}$ is an estimator of X based on Y.
- Let $h(Y)$ be any other estimator of X based on Y.
- We just showed that, for every value of y, we have:

  $E[(X-g(y))^2|Y=y] \leq E[(X-h(y))^2|Y=y]$
- Therefore,

  $E[(X-g(Y))^2|Y] \leq E[(X-h(Y))^2|Y]$
- Take expectations of both sides and use the iterated expectation formula:

  $E[(X-g(Y))^2] \leq E[(X-h(Y))^2]$
- Therefore, the conditional expectation estimator achieves an MSE which is ≤ to the MSE for any other estimator.

Other names that LMS estimate goes by in literature

- Least-squares estimate (LSE)
- Bayes least-squares estimate (BLSE)
- Minimum mean square error estimate (MMSE)

Ex 8.3: estimating the mean of a Gaussian r.v.

- Observe $Y_i = X + W_i$ for $i = 1,\ldots,n$
- $X, W_1, \ldots, W_n$ independent Gaussian, with known means and variances
- W_i has mean 0 and variance σ_i^2
- X has mean x_0 and variance σ_0^2
- Previously, we found the MAP estimate of X based on observing $\mathbf{Y} = (Y_1, \ldots, Y_n)$
- Now let's find the LMS estimate of X based on observing $\mathbf{Y} = (Y_1, \ldots, Y_n)$

Prior model and measurement model

Prior model:

$$f_X(x) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x_0}{\sigma_0}\right)^2}$$

Measurement model:

$$f_{\mathbf{Y}|X}(\mathbf{y} \mid x) = \prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y_i-x}{\sigma_i}\right)^2}$$

$$= (2\pi)^{-n/2}\left(\prod_{i=1}^{n}\frac{1}{\sigma_i}\right)\exp\left(-\sum_{i=1}^{n}\frac{1}{2}\left(\frac{y_i-x}{\sigma_i}\right)^2\right)$$

Posterior PDF

$$f_{X|\mathbf{Y}}(x \mid \mathbf{y}) = \frac{f_{\mathbf{Y}|X}(\mathbf{y} \mid x) f_X(x)}{f_{\mathbf{Y}}(\mathbf{y})}$$

Posterior PDF

$$f_X(x) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x_0}{\sigma_0}\right)^2}$$

$$f_{X|\mathbf{Y}}(x \mid \mathbf{y}) = \frac{f_{\mathbf{Y}|X}(\mathbf{y} \mid x) f_X(x)}{f_{\mathbf{Y}}(\mathbf{y})}$$

Posterior PDF

$$f_X(x) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x_0}{\sigma_0}\right)^2}$$

$$f_{\mathbf{Y}|X}(\mathbf{y}\,|\,x) = (2\pi)^{-n/2}\left(\prod_{i=1}^{n}\frac{1}{\sigma_i}\right)\exp\left(-\sum_{i=1}^{n}\frac{1}{2}\left(\frac{y_i - x}{\sigma_i}\right)^2\right)$$

$$f_{X|\mathbf{Y}}(x\,|\,\mathbf{y}) = \frac{f_{\mathbf{Y}|X}(\mathbf{y}\,|\,x)\, f_X(x)}{f_{\mathbf{Y}}(\mathbf{y})}$$

Posterior PDF

$$f_X(x) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x_0}{\sigma_0}\right)^2}$$

$$f_{\mathbf{Y}|X}(\mathbf{y} \mid x) = (2\pi)^{-n/2}\left(\prod_{i=1}^{n}\frac{1}{\sigma_i}\right)\exp\left(-\sum_{i=1}^{n}\frac{1}{2}\left(\frac{y_i - x}{\sigma_i}\right)^2\right)$$

$$f_{X|\mathbf{Y}}(x \mid \mathbf{y}) = \frac{f_{\mathbf{Y}|X}(\mathbf{y} \mid x) f_X(x)}{f_{\mathbf{Y}}(\mathbf{y})}$$

$$= \frac{1}{f_{\mathbf{Y}}(\mathbf{y})}(2\pi)^{-n/2}\left(\prod_{i=1}^{n}\frac{1}{\sigma_i}\right)\exp\left(-\sum_{i=1}^{n}\frac{1}{2}\left(\frac{y_i - x}{\sigma_i}\right)^2\right)\frac{1}{\sigma_0\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x - x_0}{\sigma_0}\right)^2\right)$$

Posterior PDF

$$f_{X|\mathbf{Y}}(x\mid\mathbf{y})=\frac{f_{\mathbf{Y}|X}(\mathbf{y}\mid x)f_X(x)}{f_{\mathbf{Y}}(\mathbf{y})}$$

$$=\frac{1}{f_{\mathbf{Y}}(\mathbf{y})}(2\pi)^{-n/2}\left(\prod_{i=1}^{n}\frac{1}{\sigma_i}\right)\exp\left(-\sum_{i=1}^{n}\frac{1}{2}\left(\frac{y_i-x}{\sigma_i}\right)^2\right)\frac{1}{\sigma_0\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-x_0}{\sigma_0}\right)^2\right)$$

$$=C\exp\left(-\sum_{i=0}^{n}\frac{1}{2}\left(\frac{y_i-x}{\sigma_i}\right)^2\right)$$

(Denoting $y_0 = x_0$ and $C =$ the factor that's not a function of x.)

Further massaging the posterior PDF…

$$f_{X|\mathbf{Y}}(x \mid \mathbf{y}) = C \exp\left(-\sum_{i=0}^{n} \frac{1}{2} \left(\frac{y_i - x}{\sigma_i} \right)^2 \right) = C \exp\left(-\frac{1}{2} \left[x^2 \sum_{i=0}^{n} \frac{1}{\sigma_i^2} - 2x \sum_{i=0}^{n} \frac{y_i}{\sigma_i^2} + \sum_{i=0}^{n} \frac{y_i^2}{\sigma_i^2} \right] \right)$$

Further massaging the posterior PDF…

$$f_{X|\mathbf{Y}}(x \mid \mathbf{y}) = C \exp\left(-\sum_{i=0}^{n} \frac{1}{2} \left(\frac{y_i - x}{\sigma_i} \right)^2 \right) = C \exp\left(-\frac{1}{2} \left[x^2 \sum_{i=0}^{n} \frac{1}{\sigma_i^2} - 2x \sum_{i=0}^{n} \frac{y_i}{\sigma_i^2} + \sum_{i=0}^{n} \frac{y_i^2}{\sigma_i^2} \right] \right)$$

Denoting $v = \dfrac{1}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$ and $m = \dfrac{\sum_{i=0}^{n} \frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$, we have:

Further massaging the posterior PDF…

$$f_{X|\mathbf{Y}}(x\,|\,\mathbf{y}) = C\exp\left(-\sum_{i=0}^{n}\frac{1}{2}\left(\frac{y_i - x}{\sigma_i}\right)^2\right) = C\exp\left(-\frac{1}{2}\left[x^2\sum_{i=0}^{n}\frac{1}{\sigma_i^2} - 2x\sum_{i=0}^{n}\frac{y_i}{\sigma_i^2} + \sum_{i=0}^{n}\frac{y_i^2}{\sigma_i^2}\right]\right)$$

Denoting $v = \dfrac{1}{\sum_{i=0}^{n}\frac{1}{\sigma_i^2}}$ and $m = \dfrac{\sum_{i=0}^{n}\frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n}\frac{1}{\sigma_i^2}}$, we have:

$$f_{X|\mathbf{Y}}(x\,|\,\mathbf{y}) = C\exp\left(-\frac{1}{2v}\left[x^2 - 2xm + m^2\right] + \frac{m^2}{2v} - \frac{1}{2}\sum_{i=0}^{n}\frac{y_i^2}{\sigma_i^2}\right)$$

Further massaging the posterior PDF…

$$f_{X|\mathbf{Y}}(x\,|\,\mathbf{y}) = C\exp\left(-\sum_{i=0}^{n}\frac{1}{2}\left(\frac{y_i - x}{\sigma_i}\right)^2\right) = C\exp\left(-\frac{1}{2}\left[x^2\sum_{i=0}^{n}\frac{1}{\sigma_i^2} - 2x\sum_{i=0}^{n}\frac{y_i}{\sigma_i^2} + \sum_{i=0}^{n}\frac{y_i^2}{\sigma_i^2}\right]\right)$$

Denoting $v = \dfrac{1}{\sum_{i=0}^{n}\frac{1}{\sigma_i^2}}$ and $m = \dfrac{\sum_{i=0}^{n}\frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n}\frac{1}{\sigma_i^2}}$, we have:

$$f_{X|\mathbf{Y}}(x\,|\,\mathbf{y}) = C\exp\left(-\frac{1}{2v}\left[x^2 - 2xm + m^2\right] + \frac{m^2}{2v} - \frac{1}{2}\sum_{i=0}^{n}\frac{y_i^2}{\sigma_i^2}\right)$$

$$= C_1\exp\left(-\frac{(x-m)^2}{2v}\right)$$

Further massaging the posterior PDF…

$$f_{X|\mathbf{Y}}(x\mid\mathbf{y}) = C\exp\left(-\sum_{i=0}^{n}\frac{1}{2}\left(\frac{y_i - x}{\sigma_i}\right)^2\right) = C\exp\left(-\frac{1}{2}\left[x^2\sum_{i=0}^{n}\frac{1}{\sigma_i^2} - 2x\sum_{i=0}^{n}\frac{y_i}{\sigma_i^2} + \sum_{i=0}^{n}\frac{y_i^2}{\sigma_i^2}\right]\right)$$

Denoting $v = \dfrac{1}{\sum_{i=0}^{n}\frac{1}{\sigma_i^2}}$ and $m = \dfrac{\sum_{i=0}^{n}\frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n}\frac{1}{\sigma_i^2}}$, we have:

$$f_{X|\mathbf{Y}}(x\mid\mathbf{y}) = C\exp\left(-\frac{1}{2v}\left[x^2 - 2xm + m^2\right] + \frac{m^2}{2v} - \frac{1}{2}\sum_{i=0}^{n}\frac{y_i^2}{\sigma_i^2}\right)$$

$$= C_1\exp\left(-\frac{(x-m)^2}{2v}\right)$$

This is a Gaussian PDF with mean m and variance v,

Further massaging the posterior PDF…

$$f_{X|\mathbf{Y}}(x\,|\,\mathbf{y}) = C\exp\left(-\sum_{i=0}^{n}\frac{1}{2}\left(\frac{y_i - x}{\sigma_i}\right)^2\right) = C\exp\left(-\frac{1}{2}\left[x^2\sum_{i=0}^{n}\frac{1}{\sigma_i^2} - 2x\sum_{i=0}^{n}\frac{y_i}{\sigma_i^2} + \sum_{i=0}^{n}\frac{y_i^2}{\sigma_i^2}\right]\right)$$

Denoting $v = \dfrac{1}{\sum_{i=0}^{n}\frac{1}{\sigma_i^2}}$ and $m = \dfrac{\sum_{i=0}^{n}\frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n}\frac{1}{\sigma_i^2}}$, we have:

$$\begin{aligned} f_{X|\mathbf{Y}}(x\,|\,\mathbf{y}) &= C\exp\left(-\frac{1}{2v}\left[x^2 - 2xm + m^2\right] + \frac{m^2}{2v} - \frac{1}{2}\sum_{i=0}^{n}\frac{y_i^2}{\sigma_i^2}\right) \\ &= C_1\exp\left(-\frac{(x-m)^2}{2v}\right) \end{aligned}$$

This is a Gaussian PDF with mean m and variance v,

and therefore we immediately have: $\quad C_1 = \dfrac{1}{\sqrt{2\pi v}}$

Posterior PDF and LMS estimate

$$f_{X|\mathbf{Y}}(x \mid \mathbf{y}) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x-m)^2}{2v} \right) \qquad \text{where } v = \frac{1}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}} \text{ and } m = \frac{\sum_{i=0}^{n} \frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$$

Posterior PDF and LMS estimate

$$f_{X|\mathbf{Y}}(x \mid \mathbf{y}) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x-m)^2}{2v} \right) \qquad \text{where } v = \frac{1}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}} \text{ and } m = \frac{\sum_{i=0}^{n} \frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$$

$$\text{Therefore, the LMS estimate is } E[X \mid \mathbf{Y} = \mathbf{y}] = m = \frac{\sum_{i=0}^{n} \frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}},$$

Posterior PDF and LMS estimate

$$f_{X|\mathbf{Y}}(x\mid\mathbf{y}) = \frac{1}{\sqrt{2\pi v}}\exp\left(-\frac{(x-m)^2}{2v}\right) \qquad \text{where } v = \frac{1}{\sum_{i=0}^{n}\frac{1}{\sigma_i^2}} \text{ and } m = \frac{\sum_{i=0}^{n}\frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n}\frac{1}{\sigma_i^2}}$$

Therefore, the LMS estimate is $E[X\mid\mathbf{Y}=\mathbf{y}] = m = \dfrac{\sum_{i=0}^{n}\frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n}\frac{1}{\sigma_i^2}}$,

and the LMS estimator is $E[X\mid\mathbf{Y}] = \dfrac{\sum_{i=0}^{n}\frac{Y_i}{\sigma_i^2}}{\sum_{i=0}^{n}\frac{1}{\sigma_i^2}}$

Posterior PDF and LMS estimate

$$f_{X|\mathbf{Y}}(x \mid \mathbf{y}) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x-m)^2}{2v} \right) \qquad \text{where } v = \frac{1}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}} \text{ and } m = \frac{\sum_{i=0}^{n} \frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$$

Therefore, the LMS estimate is $E[X \mid \mathbf{Y} = \mathbf{y}] = m = \dfrac{\sum_{i=0}^{n} \frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$,

and the LMS estimator is $E[X \mid \mathbf{Y}] = \dfrac{\sum_{i=0}^{n} \frac{Y_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$

Note that, in this example, LMS = MAP

Posterior PDF and LMS estimate

$$f_{X|\mathbf{Y}}(x \mid \mathbf{y}) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x-m)^2}{2v}\right) \qquad \text{where } v = \frac{1}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}} \text{ and } m = \frac{\sum_{i=0}^{n} \frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$$

Therefore, the LMS estimate is $E[X \mid \mathbf{Y} = \mathbf{y}] = m = \dfrac{\sum_{i=0}^{n} \frac{y_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$,

and the LMS estimator is $E[X \mid \mathbf{Y}] = \dfrac{\sum_{i=0}^{n} \frac{Y_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$

Note that, in this example, LMS = MAP, because:

- the posterior density is Gaussian
- for a Gaussian density, the maximum is at the mean

Ex. 8.11

- X = continuous uniform over [4, 10].
- W = continuous uniform over [-1, 1].
- X and W are independent.

Ex. 8.11

- X = continuous uniform over [4, 10].
- W = continuous uniform over [-1, 1].
- X and W are independent.
- $Y = X + W$.

Ex. 8.11

- X = continuous uniform over [4, 10].
- W = continuous uniform over [-1, 1].
- X and W are independent.
- $Y = X + W$.
- Calculate the LMS estimate of X based on $Y=y$.

Ex. 8.11

- X = continuous uniform over [4, 10].
- W = continuous uniform over [-1, 1].
- X and W are independent.
- $Y = X + W$.
- Calculate the LMS estimate of X based on $Y=y$.
- Calculate the associated conditional MSE.

Ex. 8.11

- X = continuous uniform over [4, 10].
- W = continuous uniform over [-1, 1].
- X and W are independent.
- $Y = X + W$.
- $f_{Y|X}(y|x)$ is uniform over $[x-1, x+1]$

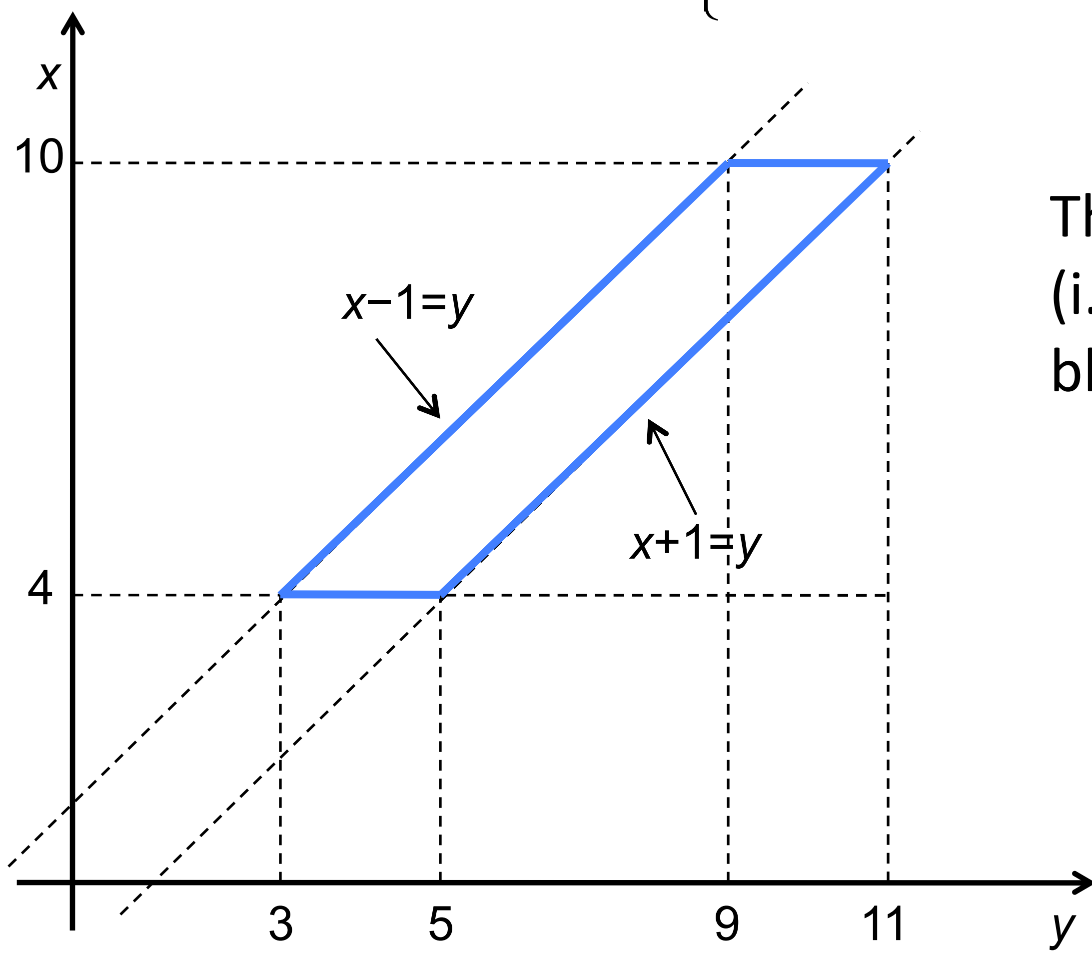
Ex. 8.11

- X = continuous uniform over [4, 10].
- W = continuous uniform over [-1, 1].
- X and W are independent.
- $Y = X + W$.
- $f_{Y|X}(y|x)$ is uniform over $[x-1, x+1]$

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y \mid x) = \begin{cases} 1/12, \text{ if } 4 \le x \le 10 \text{ and } x-1 \le y \le x+1 \\ 0, \text{ otherwise} \end{cases}$$

Ex. 8.11: support of the joint PDF of *X* and *Y*

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y\,|\,x) = \begin{cases} 1/12, \text{ if } 4 \le x \le 10 \text{ and } x-1 \le y \le x+1 \\ 0, \text{ otherwise} \end{cases}$$



The joint PDF is supported (i.e., is nonzero) inside the blue parallelogram.

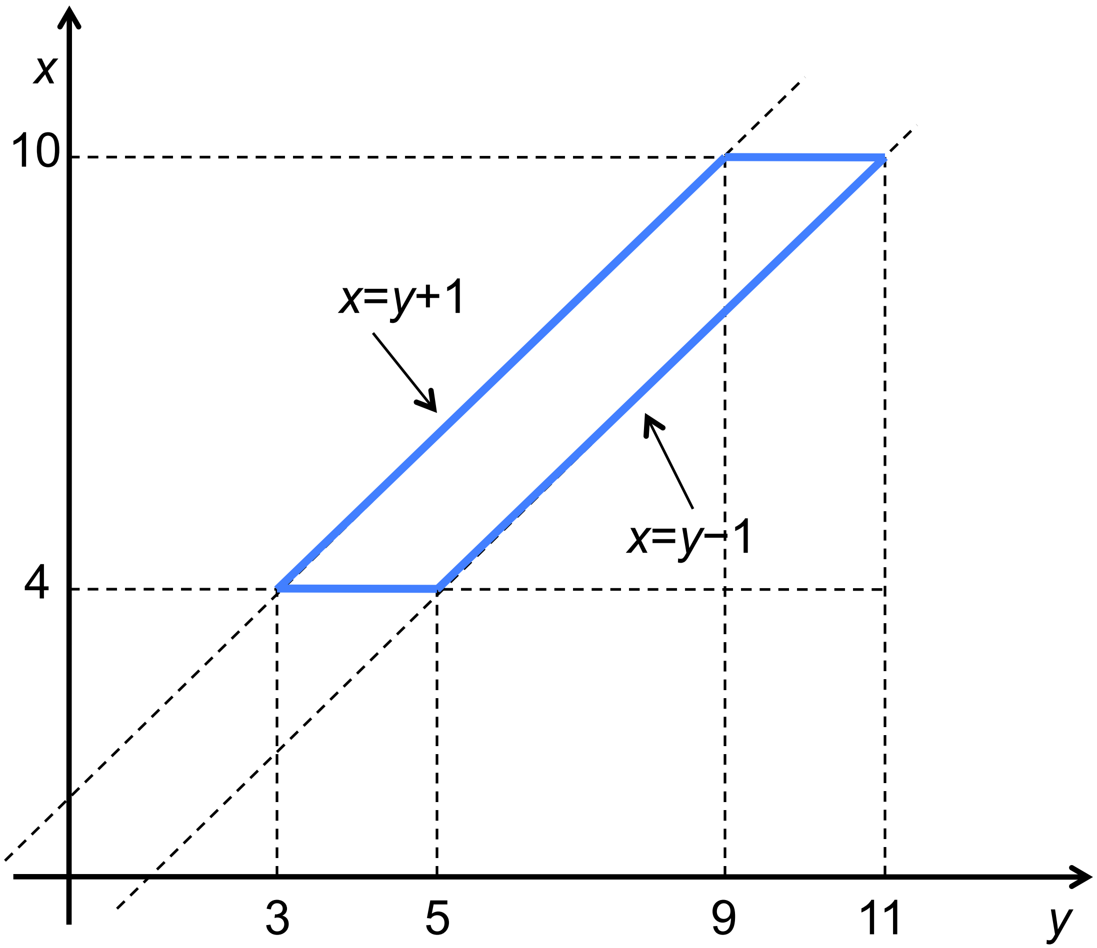
Ex. 8.11: the form of the posterior

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

Ex. 8.11: the form of the posterior

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

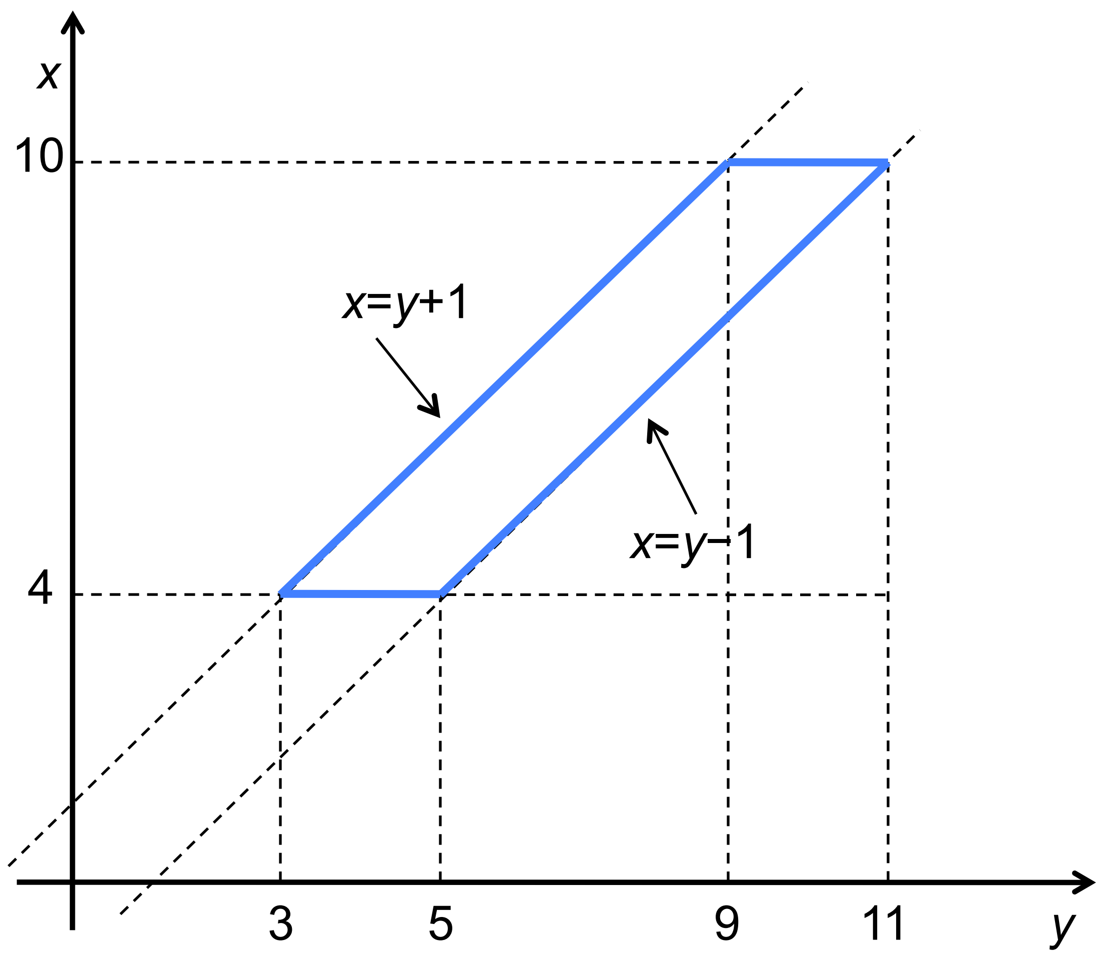
$f_{X,Y}$ is uniform over the parallelogram.



Ex. 8.11: the form of the posterior

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

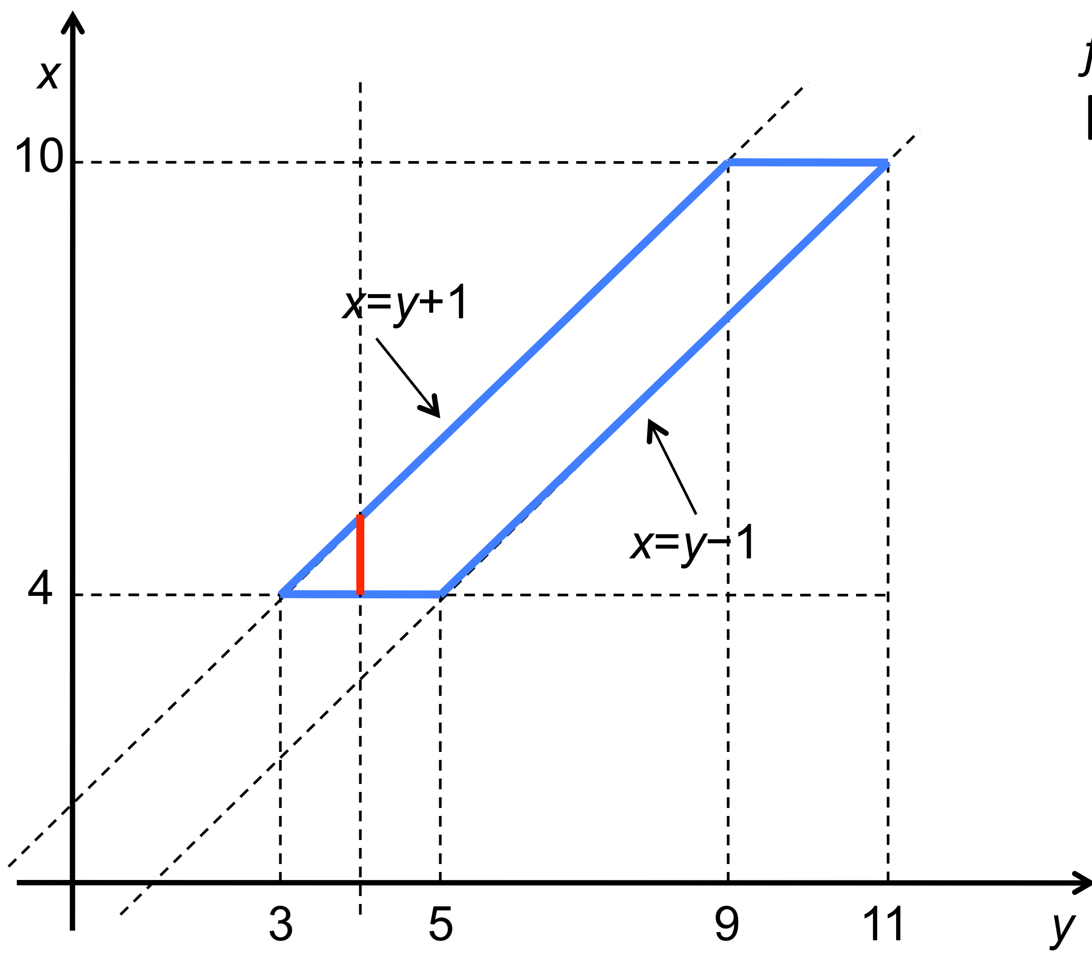
$f_{X,Y}$ is uniform over the parallelogram. Therefore, for any fixed $y=y_0$, the posterior is uniform over the intersection of the line $y=y_0$ with the parallelogram.



Ex. 8.11: the posterior

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

$f_{X,Y}$ is uniform over the parallelogram. Therefore, for any fixed $y=y_0$, the posterior is uniform over the intersection of the line $y=y_0$ with the parallelogram.



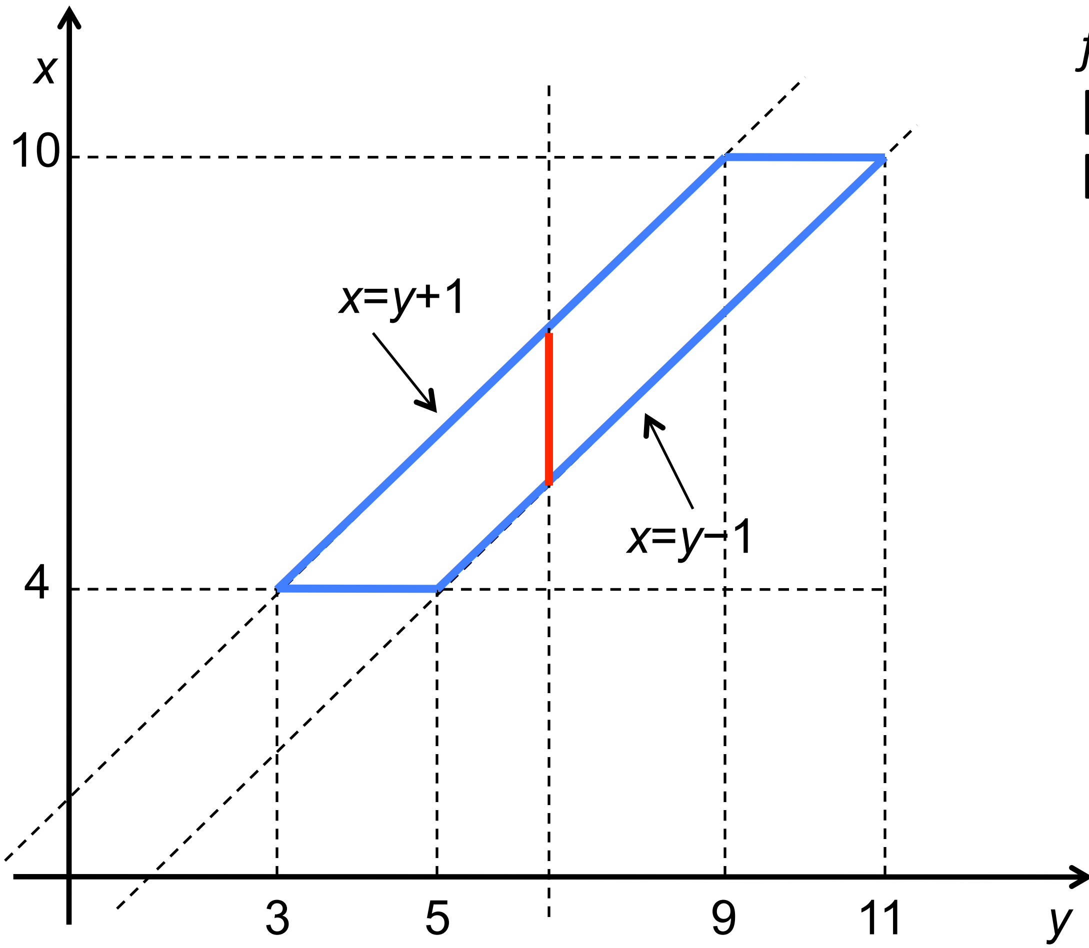
$f_{X|Y}(x|y)$ is uniform over:

$[4, y+1]$ for $3 \le y \le 5$

Ex. 8.11: the posterior

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

$f_{X,Y}$ is uniform over the parallelogram. Therefore, for any fixed $y=y_0$, the posterior is uniform over the intersection of the line $y=y_0$ with the parallelogram.



$f_{X|Y}(x|y)$ is uniform over:

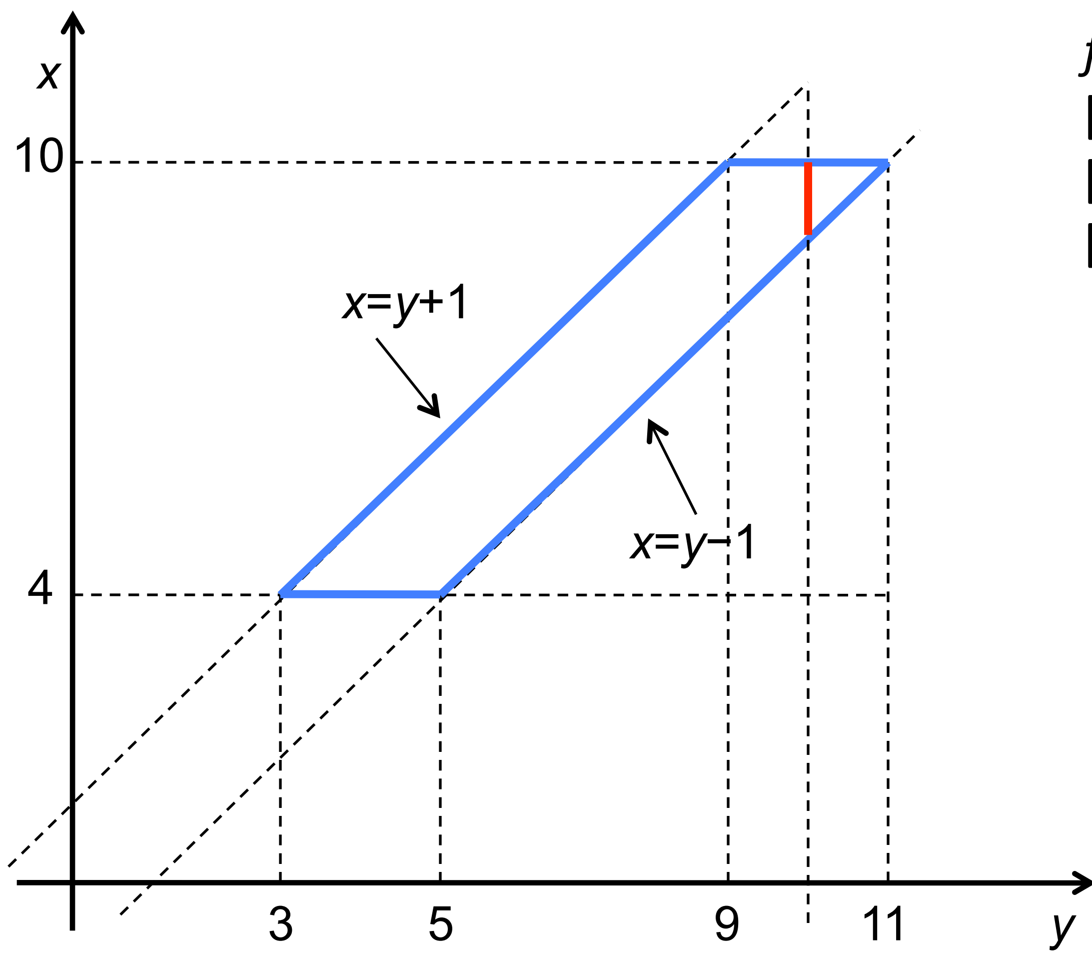
$[4, y+1]$ for $3 \le y \le 5$

$[y-1, y+1]$ for $5 \le y \le 9$

Ex. 8.11: the posterior

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

$f_{X,Y}$ is uniform over the parallelogram. Therefore, for any fixed $y=y_0$, the posterior is uniform over the intersection of the line $y=y_0$ with the parallelogram.



$f_{X|Y}(x|y)$ is uniform over:

$[4, y+1]$ for $3 \le y \le 5$

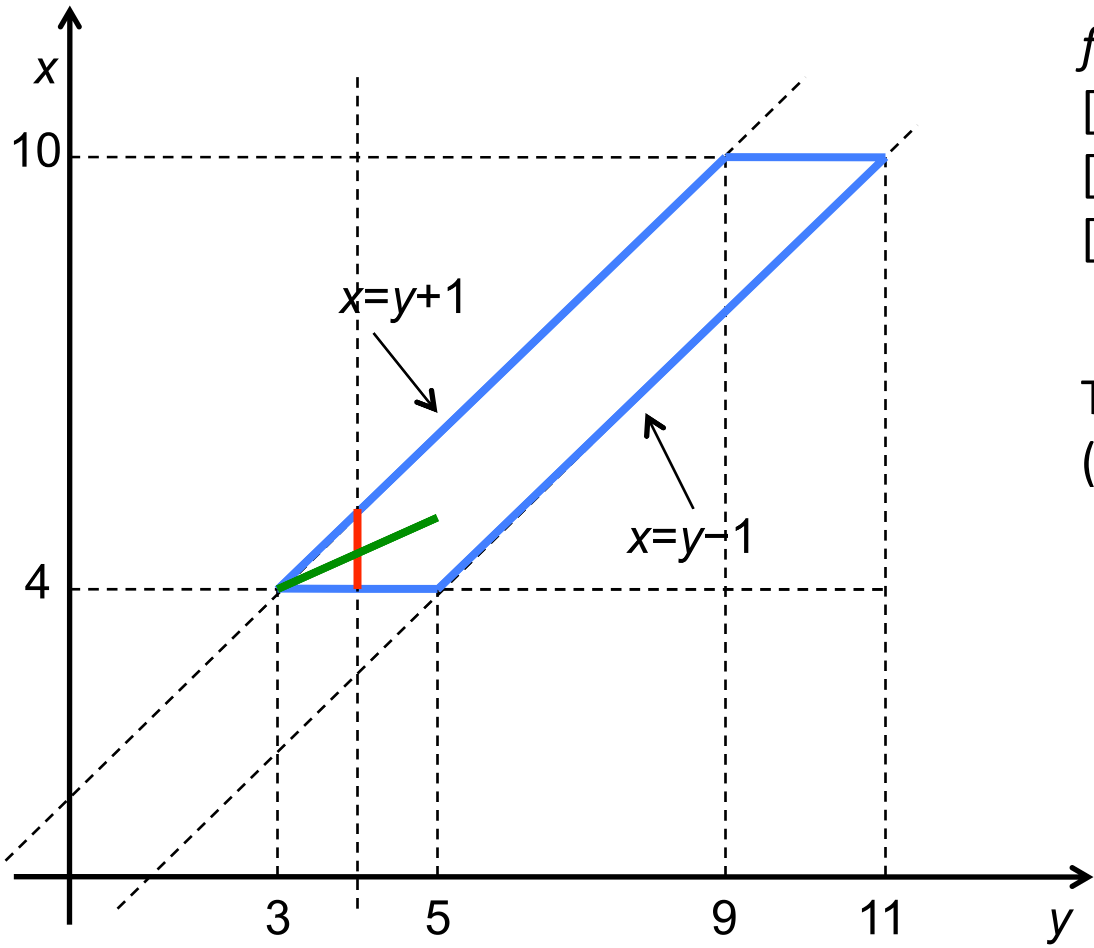
$[y-1, y+1]$ for $5 \le y \le 9$

$[y-1, 10]$ for $9 \le y \le 11$

Ex. 8.11: the LMS estimate

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

$f_{X,Y}$ is uniform over the parallelogram. Therefore, for any fixed $y=y_0$, the posterior is uniform over the intersection of the line $y=y_0$ with the parallelogram.



$f_{X|Y}(x|y)$ is uniform over:

$[4, y+1]$ for $3 \le y \le 5$

$[y-1, y+1]$ for $5 \le y \le 9$

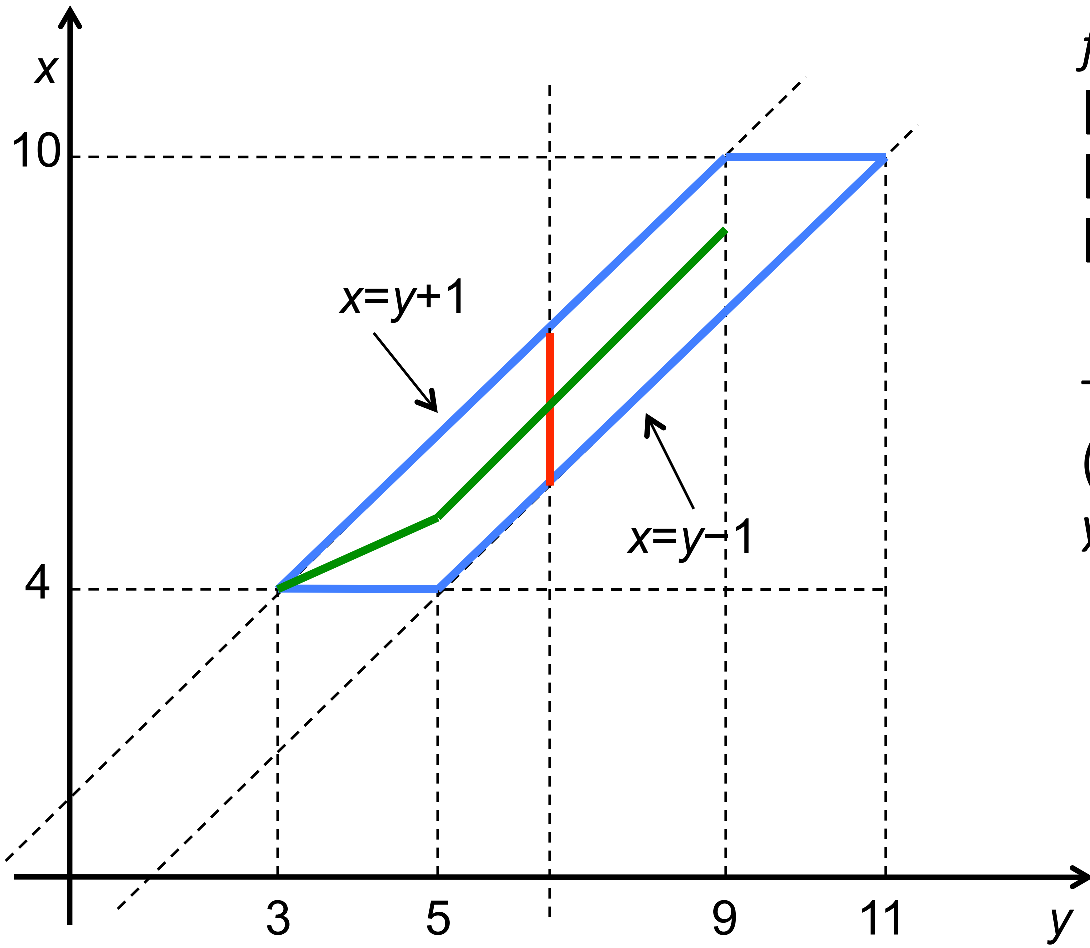
$[y-1, 10]$ for $9 \le y \le 11$

Therefore, $E[X|Y=y] = (y+5)/2$ for $3 \le y \le 5$

Ex. 8.11: the LMS estimate

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

$f_{X,Y}$ is uniform over the parallelogram. Therefore, for any fixed $y=y_0$, the posterior is uniform over the intersection of the line $y=y_0$ with the parallelogram.



$f_{X|Y}(x|y)$ is uniform over:

$[4, y+1]$ for $3 \le y \le 5$

$[y-1, y+1]$ for $5 \le y \le 9$

$[y-1, 10]$ for $9 \le y \le 11$

Therefore, $E[X|Y=y]$ =

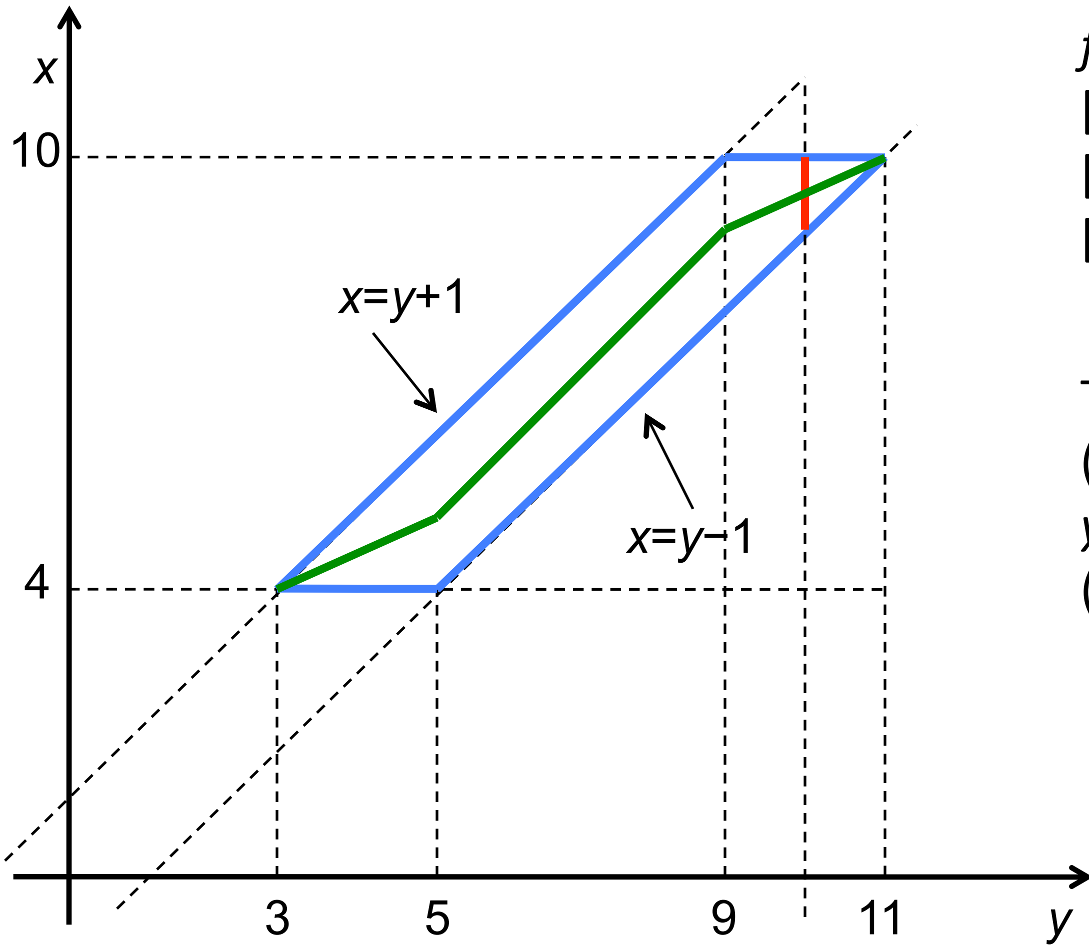
$(y+5)/2$ for $3 \le y \le 5$

y for $5 \le y \le 9$

Ex. 8.11: the LMS estimate

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

$f_{X,Y}$ is uniform over the parallelogram. Therefore, for any fixed $y=y_0$, the posterior is uniform over the intersection of the line $y=y_0$ with the parallelogram.



$f_{X|Y}(x|y)$ is uniform over:

$[4, y+1]$ for $3 \le y \le 5$

$[y-1, y+1]$ for $5 \le y \le 9$

$[y-1, 10]$ for $9 \le y \le 11$

Therefore, $E[X|Y=y]$ =

$(y+5)/2$ for $3 \le y \le 5$

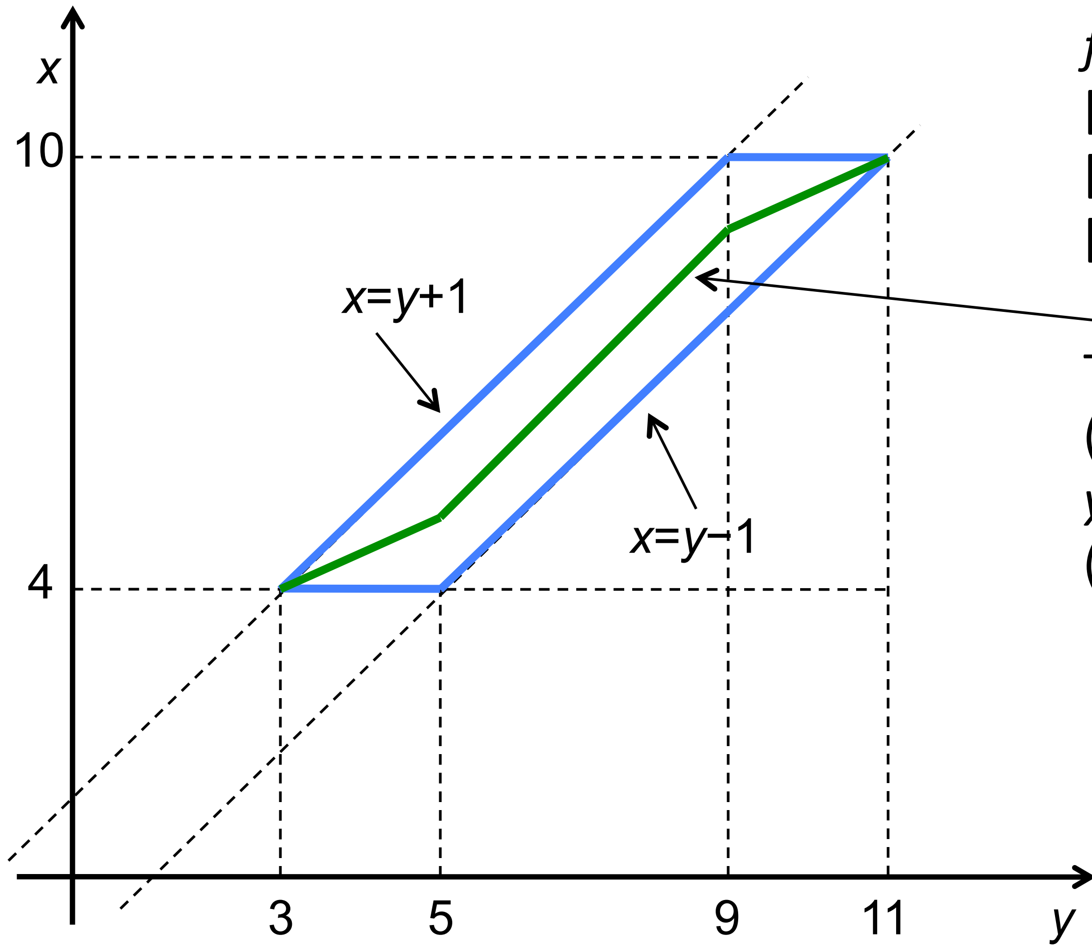
y for $5 \le y \le 9$

$(y+9)/2$ for $9 \le y \le 11$

Ex. 8.11: the LMS estimate

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

$f_{X,Y}$ is uniform over the parallelogram. Therefore, for any fixed $y=y_0$, the posterior is uniform over the intersection of the line $y=y_0$ with the parallelogram.



$f_{X|Y}(x|y)$ is uniform over:

$[4, y+1]$ for $3 \le y \le 5$

$[y-1, y+1]$ for $5 \le y \le 9$

$[y-1, 10]$ for $9 \le y \le 11$

Therefore, $E[X|Y=y]$ =

$(y+5)/2$ for $3 \le y \le 5$

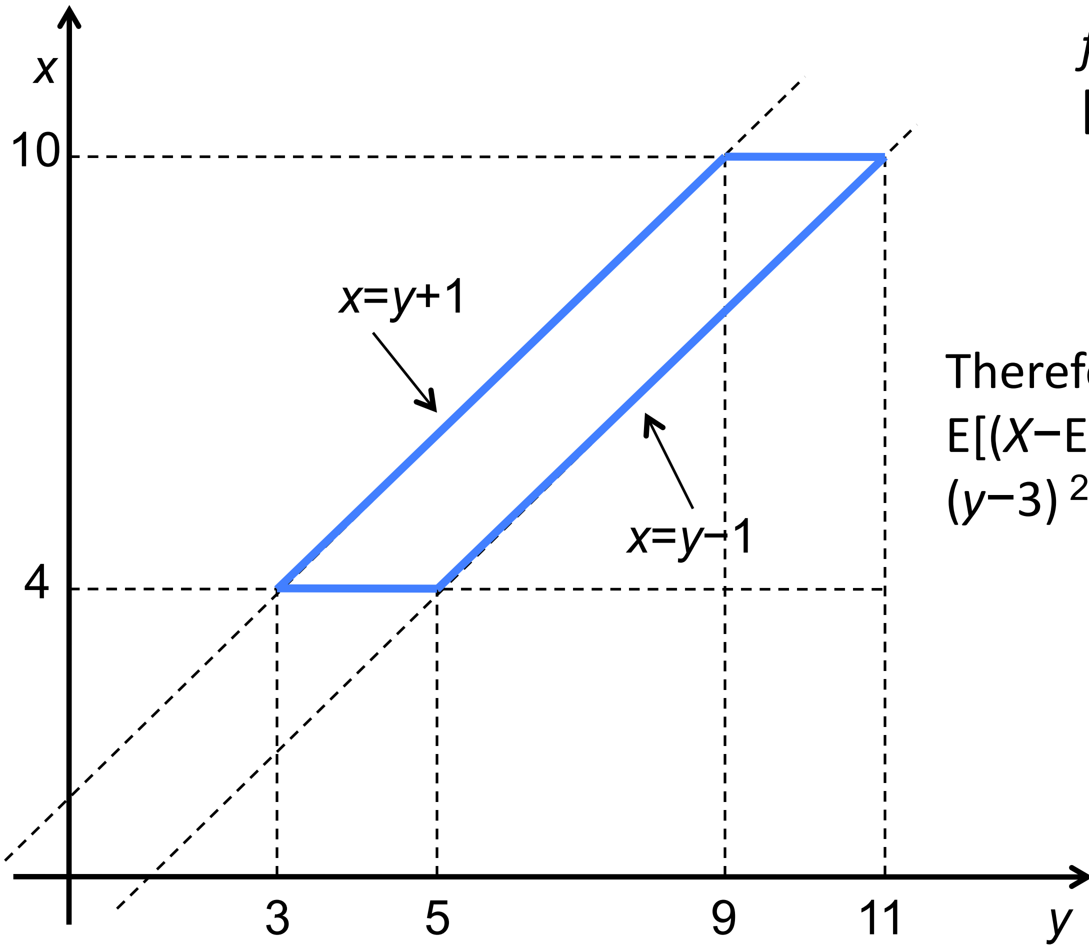
y for $5 \le y \le 9$

$(y+9)/2$ for $9 \le y \le 11$

Ex. 8.11: the conditional MSE

$$E[(X - E[X|Y=y])^2 \mid Y=y] = \mathrm{var}(X|Y=y)$$

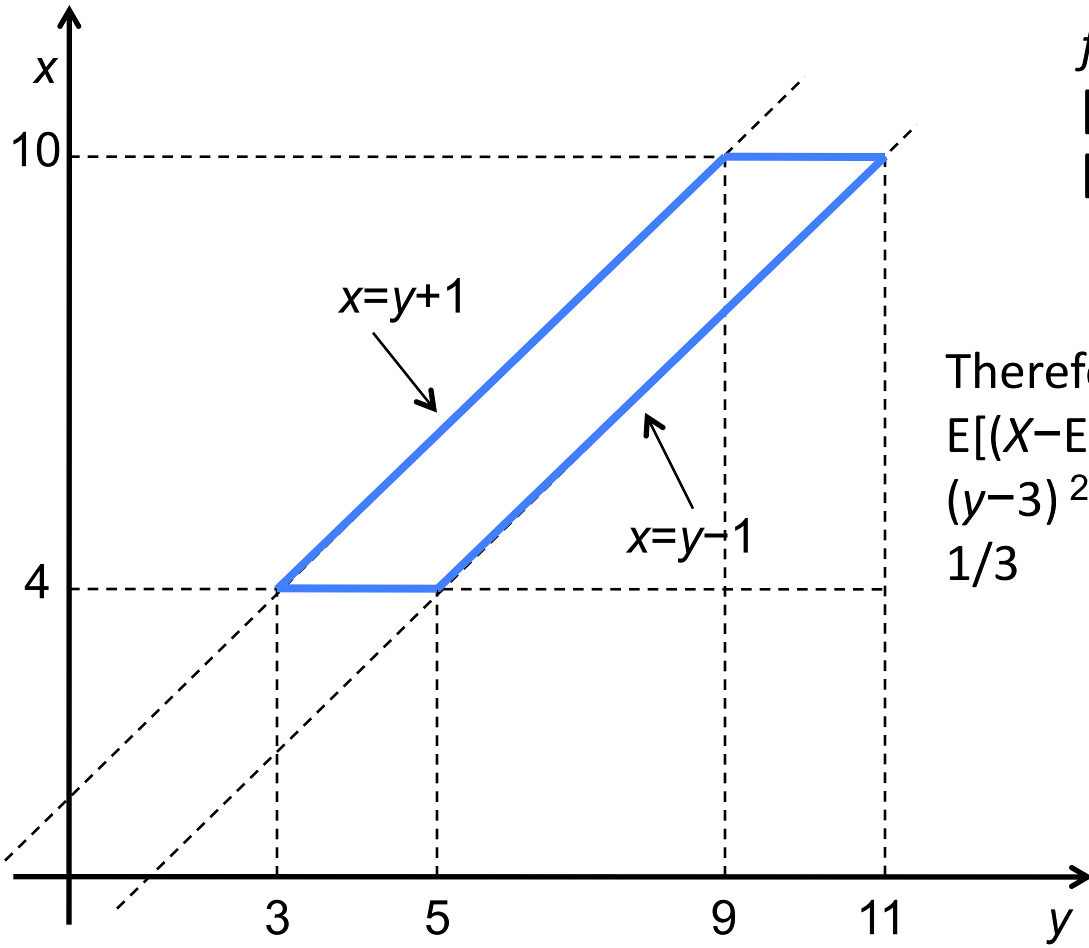
Ex. 8.11: the conditional MSE



$f_{X|Y}(x|y)$ is uniform over:
$[4, y+1]$ for $3 \le y \le 5$

Therefore,
$E[(X-E[X|Y=y])^2 \mid Y=y] = \mathrm{var}(X|Y=y) =$
$(y-3)^2/12$ for $3 \le y \le 5$

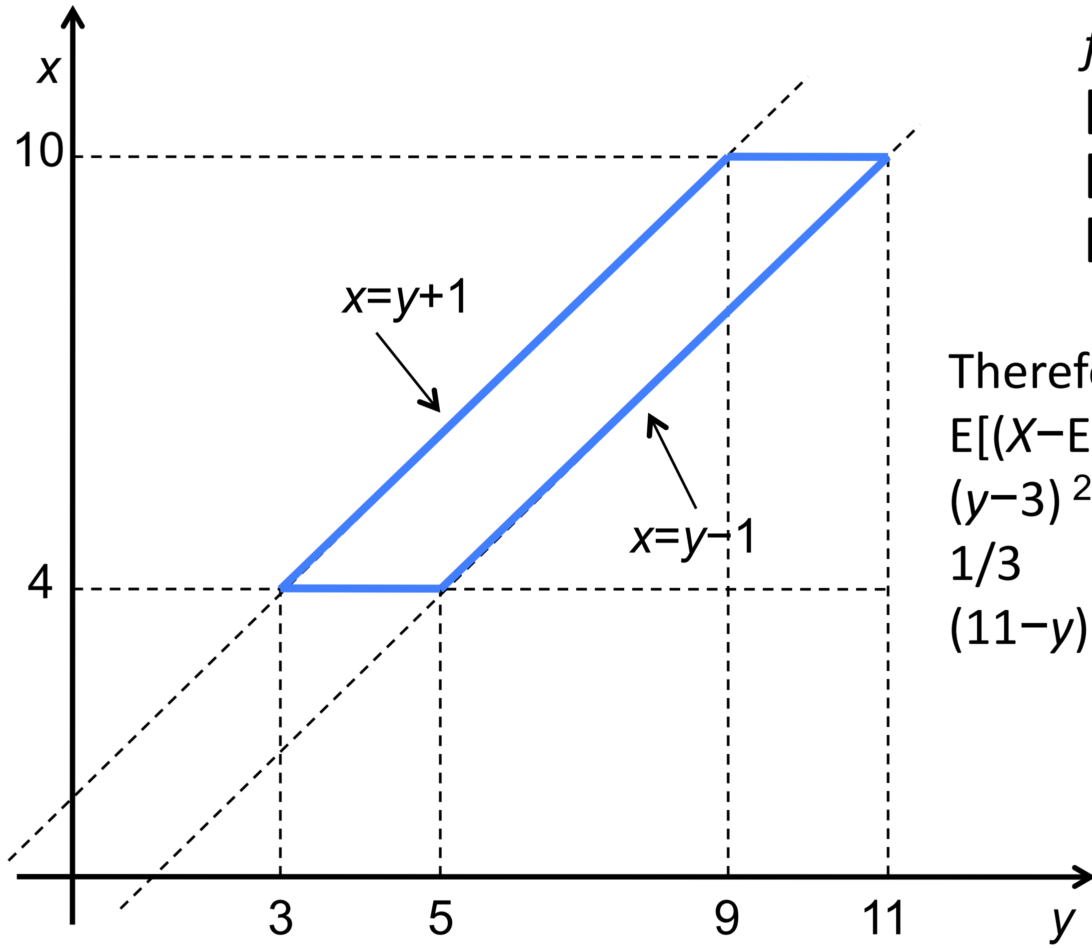
Ex. 8.11: the conditional MSE



$f_{X|Y}(x|y)$ is uniform over:
$[4, y+1]$ for $3 \le y \le 5$
$[y-1, y+1]$ for $5 \le y \le 9$

Therefore,
$E[(X - E[X|Y=y])^2 \mid Y=y] = \mathrm{var}(X|Y=y) =$
$(y-3)^2/12$ for $3 \le y \le 5$
$1/3$ for $5 \le y \le 9$

Ex. 8.11: the conditional MSE



$f_{X|Y}(x|y)$ is uniform over:

$[4, y+1]$ for $3 \le y \le 5$

$[y-1, y+1]$ for $5 \le y \le 9$

$[y-1, 10]$ for $9 \le y \le 11$

Therefore,

$E[(X-E[X|Y=y])^2 \mid Y=y] = \mathrm{var}(X|Y=y) =$

$(y-3)^2/12$ for $3 \le y \le 5$

$1/3$ for $5 \le y \le 9$

$(11-y)^2/12$ for $9 \le y \le 11$

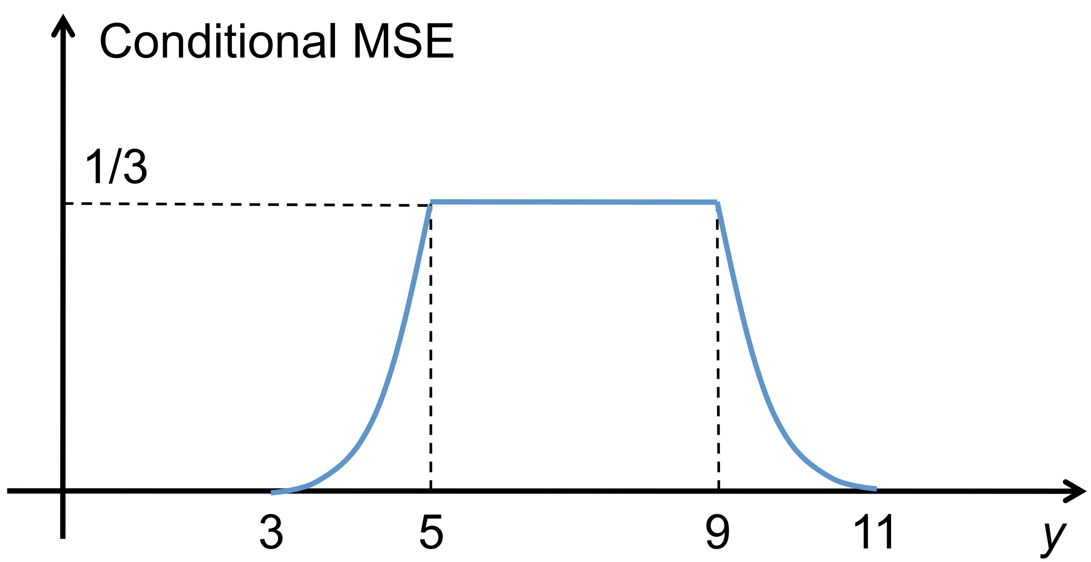
Ex. 8.11: the conditional MSE

$E[(X-E[X|Y=y])^2 \mid Y=y] = \text{var}(X|Y=y) =$

$(y-3)^2/12$ for $3 \le y \le 5$

$1/3$ for $5 \le y \le 9$

$(11-y)^2/12$ for $9 \le y \le 11$



Estimation error for LMS

We denote the LMS estimator of X based on Y by $\hat{X}_{\text{LMS}}$:

$$\hat{X}_{\text{LMS}} = E[X \mid Y]$$

The associated estimation error is denoted by $\tilde{X}_{\text{LMS}}$ and is defined as:

$$\tilde{X}_{\text{LMS}} = \hat{X}_{\text{LMS}} - X = E[X \mid Y] - X$$

Properties of the estimation error for LMS

We denote the LMS estimator of X based on Y by $\hat{X}_{\text{LMS}}$:

$$\hat{X}_{\text{LMS}} = E[X \mid Y]$$

The associated estimation error is denoted by $\tilde{X}_{\text{LMS}}$ and is defined as:

$$\tilde{X}_{\text{LMS}} = \hat{X}_{\text{LMS}} - X = E[X \mid Y] - X$$

The estimation error has zero mean:

$$E\left[\tilde{X}_{\text{LMS}}\right] = E\left[E[X \mid Y] - X\right] = E\left[E[X \mid Y]\right] - E\left[X\right] = 0$$

Properties of the estimation error for LMS

We denote the LMS estimator of X based on Y by $\hat{X}_{\text{LMS}}$:

$$\hat{X}_{\text{LMS}} = E[X \mid Y]$$

The associated estimation error is denoted by $\tilde{X}_{\text{LMS}}$ and is defined as:

$$\tilde{X}_{\text{LMS}} = \hat{X}_{\text{LMS}} - X = E[X \mid Y] - X$$

The estimation error has zero mean:

$$E\left[\tilde{X}_{\text{LMS}}\right] = E\left[E[X \mid Y] - X\right] = E\left[E[X \mid Y]\right] - E\left[X\right] = 0$$

The estimation error is uncorrelated with $\hat{X}_{\text{LMS}}$:

$$E\left[\hat{X}_{\text{LMS}}\tilde{X}_{\text{LMS}}\right] = 0$$

Properties of the estimation error for LMS

We denote the LMS estimator of X based on Y by $\hat{X}_{\text{LMS}}$:

$$\hat{X}_{\text{LMS}} = E[X \mid Y]$$

The associated estimation error is denoted by $\tilde{X}_{\text{LMS}}$ and is defined as:

$$\tilde{X}_{\text{LMS}} = \hat{X}_{\text{LMS}} - X = E[X \mid Y] - X$$

The estimation error has zero mean:

$$E\left[\tilde{X}_{\text{LMS}}\right] = E\left[E[X \mid Y] - X\right] = E\left[E[X \mid Y]\right] - E\left[X\right] = 0$$

The estimation error is uncorrelated with $\hat{X}_{\text{LMS}}$:

$$E\left[\hat{X}_{\text{LMS}}\tilde{X}_{\text{LMS}}\right] = 0$$

Therefore,

$$\begin{aligned}\operatorname{var}(X) &= \operatorname{var}\left(\hat{X}_{\text{LMS}} - (\hat{X}_{\text{LMS}} - X)\right) = \operatorname{var}\left(\hat{X}_{\text{LMS}} - \tilde{X}_{\text{LMS}}\right) = \operatorname{var}\left(\hat{X}_{\text{LMS}}\right) + \operatorname{var}\left(-\tilde{X}_{\text{LMS}}\right) \\ &= \operatorname{var}\left(\hat{X}_{\text{LMS}}\right) + \operatorname{var}\left(\tilde{X}_{\text{LMS}}\right)\end{aligned}$$