18.600: Lecture 25
Conditional expectation

Scott Sheffield

MIT
Outline

Conditional probability distributions

Conditional expectation

Interpretation and examples
Outline

Conditional probability distributions

Conditional expectation

Interpretation and examples
Recall: conditional probability distributions

- It all starts with the definition of conditional probability: \( P(A|B) = P(AB)/P(B) \).
Recall: conditional probability distributions

- It all starts with the definition of conditional probability:
  \[ P(A|B) = \frac{P(AB)}{P(B)}. \]

- If \( X \) and \( Y \) are jointly discrete random variables, we can use this to define a probability mass function for \( X \) given \( Y = y \).
Recall: conditional probability distributions

- It all starts with the definition of conditional probability: \( P(A|B) = P(AB)/P(B) \).
- If \( X \) and \( Y \) are jointly discrete random variables, we can use this to define a probability mass function for \( X \ given \ Y = y \).
- That is, we write \( p_{X|Y}(x|y) = P\{X = x|Y = y\} = \frac{p(x,y)}{p_Y(y)} \).
Recall: conditional probability distributions

- It all starts with the definition of conditional probability:
  \[ P(A|B) = \frac{P(AB)}{P(B)}. \]

- If \( X \) and \( Y \) are jointly discrete random variables, we can use this to define a probability mass function for \( X \) given \( Y = y \).

- That is, we write
  \[ p_{X|Y}(x|y) = P\{X = x|Y = y\} = \frac{p(x,y)}{p_Y(y)}. \]

- In words: first restrict sample space to pairs \((x, y)\) with given \( y \) value. Then divide the original mass function by \( p_Y(y) \) to obtain a probability mass function on the restricted space.
Recall: conditional probability distributions

- It all starts with the definition of conditional probability:
  \[ P(A|B) = \frac{P(AB)}{P(B)}. \]

- If \(X\) and \(Y\) are jointly discrete random variables, we can use this to define a probability mass function for \(X\) given \(Y = y\).
  
  That is, we write
  \[ p_{X|Y}(x|y) = P\{X = x|Y = y\} = \frac{p(x,y)}{p_Y(y)}. \]

  In words: first restrict sample space to pairs \((x, y)\) with given \(y\) value. Then divide the original mass function by \(p_Y(y)\) to obtain a probability mass function on the restricted space.

- We do something similar when \(X\) and \(Y\) are continuous random variables. In that case we write
  \[ f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}. \]
Recall: conditional probability distributions

- It all starts with the definition of conditional probability: 
  \[ P(A \mid B) = \frac{P(AB)}{P(B)}. \]

- If \( X \) and \( Y \) are jointly discrete random variables, we can use this to define a probability mass function for \( X \) given \( Y = y \).

- That is, we write 
  \[ p_{X \mid Y}(x \mid y) = P\{X = x \mid Y = y\} = \frac{p(x,y)}{p_Y(y)}. \]

- In words: first restrict sample space to pairs \((x, y)\) with given \( y \) value. Then divide the original mass function by \( p_Y(y) \) to obtain a probability mass function on the restricted space.

- We do something similar when \( X \) and \( Y \) are continuous random variables. In that case we write 
  \[ f_{X \mid Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)}. \]

- Often useful to think of sampling \((X, Y)\) as a two-stage process. First sample \( Y \) from its marginal distribution, obtain \( Y = y \) for some particular \( y \). Then sample \( X \) from its probability distribution given \( Y = y \).
Recall: conditional probability distributions

- It all starts with the definition of conditional probability:
  \[ P(A|B) = \frac{P(AB)}{P(B)}. \]

- If \( X \) and \( Y \) are jointly discrete random variables, we can use this to define a probability mass function for \( X \) given \( Y = y \).
  
  That is, we write
  \[ p_{X|Y}(x|y) = P\{X = x|Y = y\} = \frac{p(x,y)}{p_Y(y)}. \]

  In words: first restrict sample space to pairs \((x,y)\) with given \( y \) value. Then divide the original mass function by \( p_Y(y) \) to obtain a probability mass function on the restricted space.

- We do something similar when \( X \) and \( Y \) are continuous random variables. In that case we write
  \[ f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}. \]

- Often useful to think of sampling \((X, Y)\) as a two-stage process. First sample \( Y \) from its marginal distribution, obtain \( Y = y \) for some particular \( y \). Then sample \( X \) from its probability distribution given \( Y = y \).

- Marginal law of \( X \) is weighted average of conditional laws.
Let $X$ be value on one die roll, $Y$ value on second die roll, and write $Z = X + Y$. 

What is the probability distribution for $X$ given that $Y = 5$?

Answer: uniform on \{1, 2, 3, 4, 5, 6\}.

What is the probability distribution for $Z$ given that $Y = 5$?

Answer: uniform on \{6, 7, 8, 9, 10, 11\}.

What is the probability distribution for $Y$ given that $Z = 5$?

Answer: uniform on \{1, 2, 3, 4\}. 

Example
Example

- Let $X$ be value on one die roll, $Y$ value on second die roll, and write $Z = X + Y$.
- What is the probability distribution for $X$ given that $Y = 5$?
Example

- Let $X$ be value on one die roll, $Y$ value on second die roll, and write $Z = X + Y$.
- What is the probability distribution for $X$ given that $Y = 5$?
- Answer: uniform on $\{1, 2, 3, 4, 5, 6\}$.
- What is the probability distribution for $Z$ given that $Y = 5$?
- Answer: uniform on $\{6, 7, 8, 9, 10, 11\}$.
- What is the probability distribution for $Y$ given that $Z = 5$?
- Answer: uniform on $\{1, 2, 3, 4\}$.
Example

- Let $X$ be value on one die roll, $Y$ value on second die roll, and write $Z = X + Y$.
- What is the probability distribution for $X$ given that $Y = 5$?
  - Answer: uniform on $\{1, 2, 3, 4, 5, 6\}$.
- What is the probability distribution for $Z$ given that $Y = 5$?
Example

- Let $X$ be value on one die roll, $Y$ value on second die roll, and write $Z = X + Y$.
- What is the probability distribution for $X$ given that $Y = 5$?
  - Answer: uniform on $\{1, 2, 3, 4, 5, 6\}$.
- What is the probability distribution for $Z$ given that $Y = 5$?
  - Answer: uniform on $\{6, 7, 8, 9, 10, 11\}$.
Let $X$ be value on one die roll, $Y$ value on second die roll, and write $Z = X + Y$.

What is the probability distribution for $X$ given that $Y = 5$?

Answer: uniform on $\{1, 2, 3, 4, 5, 6\}$.

What is the probability distribution for $Z$ given that $Y = 5$?

Answer: uniform on $\{6, 7, 8, 9, 10, 11\}$.

What is the probability distribution for $Y$ given that $Z = 5$?
Let $X$ be value on one die roll, $Y$ value on second die roll, and write $Z = X + Y$.

What is the probability distribution for $X$ given that $Y = 5$?
Answer: uniform on $\{1, 2, 3, 4, 5, 6\}$.

What is the probability distribution for $Z$ given that $Y = 5$?
Answer: uniform on $\{6, 7, 8, 9, 10, 11\}$.

What is the probability distribution for $Y$ given that $Z = 5$?
Answer: uniform on $\{1, 2, 3, 4\}$. 
Conditional probability distributions

Conditional expectation

Interpretation and examples
Outline

Conditional probability distributions

Conditional expectation

Interpretation and examples
Now, what do we mean by $E[X|Y = y]$? This should just be the expectation of $X$ in the conditional probability measure for $X$ given that $Y = y$. 

Can make sense of this in the continuum setting as well. In continuum setting we had $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$. So $E[X|Y = y] = \int_{\mathbb{R}} x f_{X|Y}(x|y) f_Y(y) dx$. 
Now, what do we mean by $E[X|Y = y]$? This should just be the expectation of $X$ in the conditional probability measure for $X$ given that $Y = y$.

Can write this as

$$E[X|Y = y] = \sum_x xP\{X = x|Y = y\} = \sum_x xp_{X|Y}(x|y).$$
Now, what do we mean by $E[X|Y = y]$? This should just be the expectation of $X$ in the conditional probability measure for $X$ given that $Y = y$.

Can write this as

$$E[X|Y = y] = \sum_x xP\{X = x|Y = y\} = \sum_x xp_{X|Y}(x|y).$$

Can make sense of this in the continuum setting as well.
Now, what do we mean by $E[X|Y = y]$? This should just be the expectation of $X$ in the conditional probability measure for $X$ given that $Y = y$.

Can write this as

$$E[X|Y = y] = \sum_x x P\{X = x|Y = y\} = \sum_x x p_{X|Y}(x|y).$$

Can make sense of this in the continuum setting as well.

In continuum setting we had $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$. So

$$E[X|Y = y] = \int_{-\infty}^{\infty} x \frac{f(x,y)}{f_Y(y)} dx$$
Example

Let $X$ be value on one die roll, $Y$ value on second die roll, and write $Z = X + Y$. 
Let $X$ be value on one die roll, $Y$ value on second die roll, and write $Z = X + Y$.

What is $E[X|Y = 5]$?
Let $X$ be value on one die roll, $Y$ value on second die roll, and write $Z = X + Y$.

What is $E[X|Y = 5]$?

What is $E[Z|Y = 5]$?
Example

- Let $X$ be value on one die roll, $Y$ value on second die roll, and write $Z = X + Y$.
- What is $E[X|Y = 5]$?
- What is $E[Z|Y = 5]$?
- What is $E[Y|Z = 5]$?
Can think of $E[X|Y]$ as a function of the random variable $Y$. When $Y = y$ it takes the value $E[X|Y = y]$. 

Very useful fact: $E[E[X|Y]] = E[X]$. In words: what you expect to expect $X$ to be after learning $Y$ is the same as what you now expect $X$ to be.
Conditional expectation as a random variable

- Can think of $E[X|Y]$ as a function of the random variable $Y$. When $Y = y$ it takes the value $E[X|Y = y]$.

- So $E[X|Y]$ is itself a random variable. It happens to depend only on the value of $Y$. 

Proof in discrete case:

$$E[E[X|Y = y]] = \sum_y p_Y(y) \sum_x x p\{X=x|Y=y\} = \sum_x x \sum_y p\{X=x,Y=y\} p_Y(y) = E[X].$$

Recall that, in general, $E[g(Y)] = \sum_y p_Y(y) g(y)$. 

$$E[E[X|Y = y]] = \sum_y p_Y(y) \sum_x x p\{X=x|Y=y\} = \sum_x x \sum_y p\{X=x,Y=y\} p_Y(y) = E[X].$$
Can think of $E[X|Y]$ as a function of the random variable $Y$. When $Y = y$ it takes the value $E[X|Y = y]$.

So $E[X|Y]$ is itself a random variable. It happens to depend only on the value of $Y$.

Thinking of $E[X|Y]$ as a random variable, we can ask what its expectation is. What is $E[E[X|Y]]$?
Conditional expectation as a random variable

- Can think of $E[X|Y]$ as a function of the random variable $Y$. When $Y = y$ it takes the value $E[X|Y = y]$.
- So $E[X|Y]$ is itself a random variable. It happens to depend only on the value of $Y$.
- Thinking of $E[X|Y]$ as a random variable, we can ask what its expectation is. What is $E[E[X|Y]]$?
- **Very useful fact:** $E[E[X|Y]] = E[X]$. 
Can think of $E[X|Y]$ as a function of the random variable $Y$. When $Y = y$ it takes the value $E[X|Y = y]$.

So $E[X|Y]$ is itself a random variable. It happens to depend only on the value of $Y$.

Thinking of $E[X|Y]$ as a random variable, we can ask what its expectation is. What is $E[E[X|Y]]$?

**Very useful fact:** $E[E[X|Y]] = E[X]$.

In words: what you expect to expect $X$ to be *after learning* $Y$ is same as what you *now* expect $X$ to be.
Can think of $E[X|Y]$ as a function of the random variable $Y$. When $Y = y$ it takes the value $E[X|Y = y]$.

So $E[X|Y]$ is itself a random variable. It happens to depend only on the value of $Y$.

Thinking of $E[X|Y]$ as a random variable, we can ask what its expectation is. What is $E[E[X|Y]]$?

**Very useful fact:** $E[E[X|Y]] = E[X]$.

In words: what you expect to expect $X$ to be after learning $Y$ is same as what you now expect $X$ to be.

Proof in discrete case:

$E[X|Y = y] = \sum_x xP\{X = x|Y = y\} = \sum_x x \frac{p(x,y)}{p_Y(y)}$. 

Recall that, in general, $E[g(Y)] = \sum_y p_Y(y)g(y)$. 

$E[E[X|Y]] = \sum_y p_Y(y)E[X|Y = y] = \sum_y p_Y(y)\sum_x x \frac{p(x,y)}{p_Y(y)} = E[X]$. 

Conditional expectation as a random variable

- Can think of $E[X|Y]$ as a function of the random variable $Y$. When $Y = y$ it takes the value $E[X|Y = y]$.
- So $E[X|Y]$ is itself a random variable. It happens to depend only on the value of $Y$.
- Thinking of $E[X|Y]$ as a random variable, we can ask what its expectation is. What is $E[E[X|Y]]$?
- **Very useful fact:** $E[E[X|Y]] = E[X]$.
- In words: what you expect to expect $X$ to be after learning $Y$ is same as what you now expect $X$ to be.
- Proof in discrete case:
  
  
  
  
  
  $E[X|Y = y] = \sum_x xP\{X = x|Y = y\} = \sum_x x \frac{p(x,y)}{p_Y(y)}$.
  
  
  
  
  Recall that, in general, $E[g(Y)] = \sum_y p_Y(y)g(y)$. 

Can think of $E[X|Y]$ as a function of the random variable $Y$. When $Y = y$ it takes the value $E[X|Y = y]$.

So $E[X|Y]$ is itself a random variable. It happens to depend only on the value of $Y$.

Thinking of $E[X|Y]$ as a random variable, we can ask what its expectation is. What is $E[E[X|Y]]$?

**Very useful fact:** $E[E[X|Y]] = E[X]$.

In words: what you expect to expect $X$ to be after learning $Y$ is same as what you now expect $X$ to be.

Proof in discrete case:

$E[X|Y = y] = \sum_x xp\{X = x|Y = y\} = \sum_x x \frac{p(x,y)}{p_Y(y)}$.

Recall that, in general, $E[g(Y)] = \sum_y p_Y(y)g(y)$.

$E[E[X|Y = y]] = \sum_y p_Y(y) \sum_x x \frac{p(x,y)}{p_Y(y)} = \sum_x \sum_y p(x,y)x = E[X]$. 
Definition: 
\[ \text{Var}(X|Y) = E[(X - E[X|Y])^2|Y] = E[X^2 - E[X|Y]^2|Y]. \]
Conditional variance

- **Definition:**
  \[ \text{Var}(X|Y) = E \left[ (X - E[X|Y])^2 | Y \right] = E \left[ X^2 - E[X|Y]^2 | Y \right]. \]

- **Var**\((X|Y)\) is a random variable that depends on \(Y\). It is the variance of \(X\) in the conditional distribution for \(X\) given \(Y\).
Definition:
\[
\text{Var}(X|Y) = E [(X - E[X|Y])^2 | Y] = E [X^2 - E[X|Y]^2 | Y].
\]

Var(X|Y) is a random variable that depends on Y. It is the variance of X in the conditional distribution for X given Y.

Note
\[
\]
Definition:
\[ \text{Var}(X|Y) = E \left[ (X - E[X|Y])^2 \right| Y] = E \left[ X^2 - E[X|Y]^2 \right| Y].\]

\text{Var}(X|Y) is a random variable that depends on \( Y \). It is the variance of \( X \) in the conditional distribution for \( X \) given \( Y \).

Note \( E[\text{Var}(X|Y)] = E[E[X^2|Y]] - E[E[X|Y]^2|Y] = E[X^2] - E[E[X|Y]^2].\)

If we subtract \( E[X]^2 \) from first term and add equivalent value \( E[E[X|Y]]^2 \) to the second, RHS becomes \( \text{Var}[X] - \text{Var}[E[X|Y]] \), which implies following:
Definition:
\[ \text{Var}(X|Y) = E\left[(X - E[X|Y])^2|Y\right] = E\left[X^2 - E[X|Y]^2|Y\right]. \]

\[ \text{Var}(X|Y) \] is a random variable that depends on \( Y \). It is the variance of \( X \) in the conditional distribution for \( X \) given \( Y \).

Note \[ E[\text{Var}(X|Y)] = E[E[X^2|Y]] - E[E[X|Y]^2|Y] = E[X^2] - E[E[X|Y]^2]. \]

If we subtract \( E[X]^2 \) from first term and add equivalent value \( E[E[X|Y]]^2 \) to the second, RHS becomes \[ \text{Var}[X] - \text{Var}[E[X|Y]], \] which implies following:

**Useful fact:** \[ \text{Var}(X) = \text{Var}(E[X|Y]) + E[\text{Var}(X|Y)]. \]
Conditional variance

Definition:
\[
\text{Var}(X|Y) = E \left[ (X - E[X|Y])^2 | Y \right] = E[X^2 - E[X|Y]^2 | Y] .
\]

\text{Var}(X|Y) is a random variable that depends on } Y. \text{ It is the variance of } X \text{ in the conditional distribution for } X \text{ given } Y.


If we subtract \( E[X]^2 \) from first term and add equivalent value \( E[E[X|Y]]^2 \) to the second, RHS becomes \( \text{Var}[X] - \text{Var}[E[X|Y]] \), which implies following:

Useful fact: \( \text{Var}(X) = \text{Var}(E[X|Y]) + E[\text{Var}(X|Y)] \).

One can discover \( X \) in two stages: first sample \( Y \) from marginal and compute \( E[X|Y] \), then sample \( X \) from distribution given \( Y \) value.
Definition:
\[
\text{Var}(X|Y) = E\left[(X - E[X|Y])^2|Y\right] = E\left[X^2 - E[X|Y]^2|Y\right].
\]
\text{Var}(X|Y) is a random variable that depends on } Y. \text{ It is the variance of } X \text{ in the conditional distribution for } X \text{ given } Y.

Note \(E[\text{Var}(X|Y)] = E[E[X^2|Y]] - E[E[X|Y]^2|Y] = E[X^2] - E[E[X|Y]^2].\)

If we subtract \(E[X]^2\) from first term and add equivalent value \(E[E[X|Y]]^2\) to the second, RHS becomes \(\text{Var}[X] - \text{Var}[E[X|Y]]\), which implies following:

**Useful fact:** \(\text{Var}(X) = \text{Var}(E[X|Y]) + E[\text{Var}(X|Y)].\)

One can discover \(X\) in two stages: first sample \(Y\) from marginal and compute \(E[X|Y]\), then sample \(X\) from distribution given \(Y\) value.

Above fact breaks variance into two parts, corresponding to these two stages.
Let $X$ be a random variable of variance $\sigma_X^2$ and $Y$ an independent random variable of variance $\sigma_Y^2$ and write $Z = X + Y$. Assume $E[X] = E[Y] = 0$. 

What are the covariances $\text{Cov}(X, Y)$ and $\text{Cov}(X, Z)$? 

How about the correlation coefficients $\rho(X, Y)$ and $\rho(X, Z)$? 

What is $E[Z|X]$? And how about $\text{Var}(Z|X)$? Both of these values are functions of $X$. Former is just $X$. Latter happens to be a constant-valued function of $X$, i.e., happens not to actually depend on $X$. We have $\text{Var}(Z|X) = \sigma_Y^2$. 

Can we check the formula $\text{Var}(Z) = \text{Var}(E[Z|X]) + E[\text{Var}(Z|X)]$ in this case?
Let $X$ be a random variable of variance $\sigma_X^2$ and $Y$ an independent random variable of variance $\sigma_Y^2$ and write $Z = X + Y$. Assume $E[X] = E[Y] = 0$.

What are the covariances $\text{Cov}(X, Y)$ and $\text{Cov}(X, Z)$?

What is $E[Z | X]$? And how about $\text{Var}(Z | X)$?

Both of these values are functions of $X$. Former is just $X$. Latter happens to be a constant-valued function of $X$, i.e., it does not actually depend on $X$. We have $\text{Var}(Z | X) = \sigma_Y^2$.

Can we check the formula $\text{Var}(Z) = \text{Var}(E[Z | X]) + E[\text{Var}(Z | X)]$ in this case?
Example

Let $X$ be a random variable of variance $\sigma_X^2$ and $Y$ an independent random variable of variance $\sigma_Y^2$ and write $Z = X + Y$. Assume $E[X] = E[Y] = 0$.

What are the covariances $\text{Cov}(X, Y)$ and $\text{Cov}(X, Z)$?

How about the correlation coefficients $\rho(X, Y)$ and $\rho(X, Z)$?
Let $X$ be a random variable of variance $\sigma_X^2$ and $Y$ an independent random variable of variance $\sigma_Y^2$ and write $Z = X + Y$. Assume $E[X] = E[Y] = 0$.

What are the covariances $\text{Cov}(X, Y)$ and $\text{Cov}(X, Z)$?

How about the correlation coefficients $\rho(X, Y)$ and $\rho(X, Z)$?

What is $E[Z|X]$? And how about $\text{Var}(Z|X)$?
Let $X$ be a random variable of variance $\sigma_X^2$ and $Y$ an independent random variable of variance $\sigma_Y^2$ and write $Z = X + Y$. Assume $E[X] = E[Y] = 0$.

What are the covariances $\text{Cov}(X, Y)$ and $\text{Cov}(X, Z)$?

How about the correlation coefficients $\rho(X, Y)$ and $\rho(X, Z)$?

What is $E[Z|X]$? And how about $\text{Var}(Z|X)$?

Both of these values are functions of $X$. Former is just $X$. Latter happens to be a constant-valued function of $X$, i.e., happens not to actually depend on $X$. We have $\text{Var}(Z|X) = \sigma_Y^2$. 

Can we check the formula $\text{Var}(Z) = \text{Var}(E[Z|X]) + E[\text{Var}(Z|X)]$ in this case?
Let $X$ be a random variable of variance $\sigma^2_X$ and $Y$ an independent random variable of variance $\sigma^2_Y$ and write $Z = X + Y$. Assume $E[X] = E[Y] = 0$.

What are the covariances $\text{Cov}(X, Y)$ and $\text{Cov}(X, Z)$?

How about the correlation coefficients $\rho(X, Y)$ and $\rho(X, Z)$?

What is $E[Z|X]$? And how about $\text{Var}(Z|X)$?

Both of these values are functions of $X$. Former is just $X$. Latter happens to be a constant-valued function of $X$, i.e., happens not to actually depend on $X$. We have $\text{Var}(Z|X) = \sigma^2_Y$.

Can we check the formula $\text{Var}(Z) = \text{Var}(E[Z|X]) + E[\text{Var}(Z|X)]$ in this case?
Outline

Conditional probability distributions

Conditional expectation

Interpretation and examples
Conditional probability distributions

Conditional expectation

Interpretation and examples
Sometimes think of the expectation $E[Y]$ as a “best guess” or “best predictor” of the value of $Y$. 
Sometimes think of the expectation $E[Y]$ as a “best guess” or “best predictor” of the value of $Y$.

It is best in the sense that among all constants $m$, the expectation $E[(Y - m)^2]$ is minimized when $m = E[Y]$. 
Sometimes think of the expectation $E[Y]$ as a “best guess” or “best predictor” of the value of $Y$.

It is best in the sense that among all constants $m$, the expectation $E[(Y - m)^2]$ is minimized when $m = E[Y]$.

But what if we allow non-constant predictors? What if the predictor is allowed to depend on the value of a random variable $X$ that we can observe directly?
Sometimes think of the expectation $E[Y]$ as a “best guess” or “best predictor” of the value of $Y$.

It is best in the sense that at among all constants $m$, the expectation $E[(Y - m)^2]$ is minimized when $m = E[Y]$.

But what if we allow non-constant predictors? What if the predictor is allowed to depend on the value of a random variable $X$ that we can observe directly?

Let $g(x)$ be such a function. Then $E[(y - g(X))^2]$ is minimized when $g(X) = E[Y|X]$. 
Toss 100 coins. What’s the conditional expectation of the number of heads given that there are $k$ heads among the first fifty tosses?
Examples

- Toss 100 coins. What’s the conditional expectation of the number of heads given that there are \( k \) heads among the first fifty tosses?
- \( k + 25 \)
Examples

- Toss 100 coins. What’s the conditional expectation of the number of heads given that there are $k$ heads among the first fifty tosses?
  
  $k + 25$

- What’s the conditional expectation of the number of aces in a five-card poker hand given that the first two cards in the hand are aces?

\[\frac{2 + 3}{50}\]
Examples

- Toss 100 coins. What’s the conditional expectation of the number of heads given that there are $k$ heads among the first fifty tosses?
- $k + 25$
- What’s the conditional expectation of the number of aces in a five-card poker hand given that the first two cards in the hand are aces?
- $2 + 3 \cdot 2/50$