

Development with Refinement in Probabilistic B — Foundation and Case Study

T.S. Hoang^{1,2} Z. Jin¹ K. Robinson^{1,2} C. Morgan¹
A. McIver³

¹School of Computer Science and Engineering, University of New South Wales

²Formal Methods Research Group, National ICT Australia

³Department of Computer Science, Macquarie University

4th International Conference of B and Z Users, 2005
University of Surrey, Guildford, U.K.



Outline

- 1 Motivation
 - Extension to Probabilistic B
 - Background

- 2 Our Results/Contribution
 - Probabilistic specification substitution
 - Fundamental theorem
 - Proof Obligations for Loops
 - Case Study



Outline

1 Motivation

- Extension to Probabilistic B
- Background

2 Our Results/Contribution

- Probabilistic specification substitution
- Fundamental theorem
- Proof Obligations for Loops
- Case Study



Extending Probabilistic B

- To extend the scope of *probabilistic B* (pB) to layered developments;
- Need to introduce *probabilistic specification substitution*;
- To extend *Abstract Machine Notation* (*AMN*) to express
 - probabilistic specification substitution;
 - probabilistic invariant (*expectation*) for loops.



Extending Probabilistic B

- To extend the scope of *probabilistic B* (pB) to layered developments;
- Need to introduce *probabilistic specification substitution*;
- To extend *Abstract Machine Notation* (AMN) to express
 - probabilistic specification substitution;
 - probabilistic invariant (*expectation*) for loops.



Extending Probabilistic B

- To extend the scope of *probabilistic B* (pB) to layered developments;
- Need to introduce *probabilistic specification substitution*;
- To extend *Abstract Machine Notation* (AMN) to express
 - probabilistic specification substitution;
 - probabilistic invariant (*expectation*) for loops.



Extending Probabilistic B

- To extend the scope of *probabilistic B* (pB) to layered developments;
- Need to introduce *probabilistic specification substitution*;
- To extend *Abstract Machine Notation* (*AMN*) to express
 - probabilistic specification substitution;
 - probabilistic invariant (*expectation*) for loops.



Extending Probabilistic B

- To extend the scope of *probabilistic B* (pB) to layered developments;
- Need to introduce *probabilistic specification substitution*;
- To extend *Abstract Machine Notation* (*AMN*) to express
 - probabilistic specification substitution;
 - probabilistic invariant (*expectation*) for loops.



Changing the B-Toolkit

We have adapted the *B-Toolkit* to assist the development of pB machines. This involves:

- new syntax;
- proof obligation generation for new constructs;
- reasoning over *real* as well as Boolean.



Changing the B-Toolkit

We have adapted the *B-Toolkit* to assist the development of pB machines. This involves:

- new syntax;
- proof obligation generation for new constructs;
- reasoning over *real* as well as Boolean.



Changing the B-Toolkit

We have adapted the *B-Toolkit* to assist the development of pB machines. This involves:

- new syntax;
- proof obligation generation for new constructs;
- reasoning over *real* as well as Boolean.



Outline

1 Motivation

- Extension to Probabilistic B
- **Background**

2 Our Results/Contribution

- Probabilistic specification substitution
- Fundamental theorem
- Proof Obligations for Loops
- Case Study



Probabilistic Generalised Substitution Language

Summary

$[x := E]exp$

The expectation obtained after replacing all free occurrences of x in exp by E

$[skip]exp$

exp

$[prog_1 \text{ } p \oplus \text{ } prog_2]exp$

$p \times [prog_1]exp$
 $+ (1-p) \times [prog_2]exp$

$prog_1 \sqsubseteq prog_2$

$[prog_1]exp \Rightarrow [prog_2]exp$

$[prog_1 \parallel prog_2]exp$

$[prog_1]exp \min [prog_2]exp$

$[@y \cdot pred \Rightarrow prog]exp$

$\min(y) \cdot (pred \mid [prog]exp)$



Probabilistic Generalised Substitution Language

Summary

$[x := E]exp$

The expectation obtained after replacing all free occurrences of x in exp by E

$[skip]exp$

exp

$[prog_1 \text{ } p \oplus \text{ } prog_2]exp$

$p \times [prog_1]exp$
 $+ (1-p) \times [prog_2]exp$

$prog_1 \sqsubseteq prog_2$

$[prog_1]exp \Rightarrow [prog_2]exp$

$[prog_1 \parallel prog_2]exp$

$[prog_1]exp \min [prog_2]exp$

$[@y \cdot pred \Rightarrow prog]exp$

$\min(y) \cdot (pred \mid [prog]exp)$



Probabilistic Generalised Substitution Language

Summary

 $[x := E]exp$

The expectation obtained after replacing all free occurrences of x in exp by E

 $[skip]exp$
 exp
 $[prog_1 \text{ } p \oplus \text{ } prog_2]exp$

$$p \times [prog_1]exp + (1-p) \times [prog_2]exp$$
 $prog_1 \sqsubseteq prog_2$
 $[prog_1]exp \Rightarrow [prog_2]exp$
 $[prog_1 \parallel prog_2]exp$
 $[prog_1]exp \min [prog_2]exp$
 $[@y \cdot pred \Rightarrow prog]exp$
 $\min(y) \cdot (pred \mid [prog]exp)$


Probabilistic Generalised Substitution Language

Summary

$[x := E]exp$

The expectation obtained after replacing all free occurrences of x in exp by E

$[skip]exp$

exp

$[prog_1 \text{ } p \oplus \text{ } prog_2]exp$

$p \times [prog_1]exp$
 $+ (1-p) \times [prog_2]exp$

$prog_1 \sqsubseteq prog_2$

$[prog_1]exp \Rightarrow [prog_2]exp$

$[prog_1 \parallel prog_2]exp$

$[prog_1]exp \min [prog_2]exp$

$[@y \cdot pred \Rightarrow prog]exp$

$\min(y) \cdot (pred \mid [prog]exp)$



How *pGSL* extends *GSL*

Expectations replace predicates

Predicates (functions from state to Boolean) are widened to *Expectations* (functions from state to real).

- For consistency with Boolean logic, we use **embedded predicates**, $\langle \text{false} \rangle = 0$, and $\langle \text{true} \rangle = 1$.
- Notationally, we have kept predicates as much as possible.



How *pGSL* extends *GSL*

Expectations replace predicates

Predicates (functions from state to Boolean) are widened to *Expectations* (functions from state to real).

- For consistency with Boolean logic, we use **embedded predicates**, $\langle \text{false} \rangle = 0$, and $\langle \text{true} \rangle = 1$.
- Notationally, we have kept predicates as much as possible.



How *pGSL* extends *GSL*

Expectations replace predicates

Predicates (functions from state to Boolean) are widened to *Expectations* (functions from state to real).

- For consistency with Boolean logic, we use **embedded predicates**, $\langle \text{false} \rangle = 0$, and $\langle \text{true} \rangle = 1$.
- Notationally, we have kept predicates as much as possible.



Outline

1 Motivation

- Extension to Probabilistic B
- Background

2 Our Results/Contribution

- Probabilistic specification substitution
- Fundamental theorem
- Proof Obligations for Loops
- Case Study



Syntax

We want to have a definition in the probabilistic world which is similar to the precondition, postcondition pair.

Standard substitution

$v : \{P, Q\}$, where P and Q are predicates over the state, x .

- $v \subseteq x$
- Q can refer to the original state by using subscripted variables x_0 .

Probabilistic substitution

$v : \{A, B\}$, where A and B are **expectations over state**.

The expected value of B over the set of final distributions is at least the expected value of A over the initial distribution.



Syntax

We want to have a definition in the probabilistic world which is similar to the precondition, postcondition pair.

Standard substitution

$v : \{P, Q\}$, where P and Q are predicates over the state, x .

- $v \subseteq x$
- Q can refer to the original state by using subscripted variables x_0 .

Probabilistic substitution

$v : \{A, B\}$, where A and B are expectations over state.

The expected value of B over the set of final distributions is at least the expected value of A over the initial distribution.



Syntax

We want to have a definition in the probabilistic world which is similar to the precondition, postcondition pair.

Standard substitution

$v : \{P, Q\}$, where P and Q are predicates over the state, x .

- $v \subseteq x$
- Q can refer to the original state by using subscripted variables x_0 .

Probabilistic substitution

$v : \{A, B\}$, where A and B are **expectations over state**.

The expected value of B over the set of final distributions is at least the expected value of A over the initial distribution.



Semantics

Program

Specification $v : \{A, B\}$.

Post-expectation

Arbitrary expectation C .

Questions?

What is $[v : \{A, B\}] C$?

Semantics of probabilistic substitution

$$[v : \{A, B\}] C \hat{=} A \times [x_0 : = x] \left(\prod x \cdot \left(\frac{C}{B \times \langle w = w_0 \rangle} \right) \right)$$

(with w is the set of unchanged variables, i.e. $x - v$).

(Similar work can be seen in White[1996] and Ying[2003])



Semantics

Program

Specification $v : \{A, B\}$.

Post-expectation

Arbitrary expectation C .

Questions?

What is $[v : \{A, B\}] C$?

Semantics of probabilistic substitution

$$[v : \{A, B\}] C \cong A \times [x_0 : = x] \left(\prod x \cdot \left(\frac{C}{B \times \langle w = w_0 \rangle} \right) \right)$$

(with w is the set of unchanged variables, i.e. $x - v$).

(Similar work can be seen in White[1996] and Ying[2003])



Semantics

Program

Specification $v : \{A, B\}$.

Post-expectation

Arbitrary expectation C .

Questions?

What is $[v : \{A, B\}] C$?

Semantics of probabilistic substitution

$$[v : \{A, B\}] C \cong A \times [x_0 : = x] \left(\prod x \cdot \left(\frac{C}{B \times \langle w = w_0 \rangle} \right) \right)$$

(with w is the set of unchanged variables, i.e. $x - v$).

(Similar work can be seen in White[1996] and Ying[2003])



Semantics

Program

Specification $v : \{A, B\}$.

Post-expectation

Arbitrary expectation C .

Questions?

What is $[v : \{A, B\}] C$?

Semantics of probabilistic substitution

$$[v : \{A, B\}] C \hat{=} A \times [x_0 : = x] \left(\prod x \cdot \left(\frac{C}{B \times \langle w = w_0 \rangle} \right) \right)$$

(with w is the set of unchanged variables, i.e. $x - v$).

(Similar work can be seen in White[1996] and Ying[2003])



Example

Program $prog_1$

$$prog_1 \hat{=} c : \left\{ \frac{1}{2}, \langle c = H \rangle \right\} .$$

Post-expectation $\langle c = H \rangle$

$$\begin{aligned} & [c : \left\{ \frac{1}{2}, \langle c = H \rangle \right\}] \langle c = H \rangle \\ \equiv & \frac{1}{2} * [c_0 := c] \left(\prod c \cdot \left(\frac{\langle c = H \rangle}{\langle c = H \rangle} \right) \right) \\ \equiv & \frac{1}{2} * [c_0 := c] 1 \\ \equiv & \frac{1}{2} \end{aligned}$$



Example

Program $prog_1$

$$prog_1 \hat{=} c : \left\{ \frac{1}{2}, \langle c = H \rangle \right\} .$$

Post-expectation $\langle c = H \rangle$

$$\begin{aligned} & [c : \left\{ \frac{1}{2}, \langle c = H \rangle \right\}] \langle c = H \rangle \\ \equiv & \frac{1}{2} * [c_0 := c] \left(\prod c \cdot \left(\frac{\langle c = H \rangle}{\langle c = H \rangle} \right) \right) \\ \equiv & \frac{1}{2} * [c_0 := c] 1 \\ \equiv & \frac{1}{2} \end{aligned}$$



Outline

1 Motivation

- Extension to Probabilistic B
- Background

2 Our Results/Contribution

- Probabilistic specification substitution
- **Fundamental theorem**
- Proof Obligations for Loops
- Case Study



Theorem

Standard Theorem

Assume that $\text{prog}_1 \hat{=} v : \{P, Q\}$.

For any program prog_2 , $\text{prog}_1 \sqsubseteq \text{prog}_2$ if and only if

$$P \implies [x_0 := x] [\text{prog}_2] Q^w,$$

where $Q^w \hat{=} Q \wedge w = w_0$.

Probabilistic Theorem

Assume that $\text{prog}_1 \hat{=} v : \{A, B\}$.

For any program prog_2 , $\text{prog}_1 \sqsubseteq \text{prog}_2$ if and only if

$$A \Rightarrow [x_0 := x] [\text{prog}_2] B^w,$$

where $B^w \hat{=} B \times \langle w = w_0 \rangle$.



Theorem

Standard Theorem

Assume that $\text{prog}_1 \hat{=} v : \{P, Q\}$.

For any program prog_2 , $\text{prog}_1 \sqsubseteq \text{prog}_2$ if and only if

$$P \implies [x_0 := x] [\text{prog}_2] Q^w,$$

where $Q^w \hat{=} Q \wedge w = w_0$.

Probabilistic Theorem

Assume that $\text{prog}_1 \hat{=} v : \{A, B\}$.

For any program prog_2 , $\text{prog}_1 \sqsubseteq \text{prog}_2$ if and only if

$$A \implies [x_0 := x] [\text{prog}_2] B^w,$$

where $B^w \hat{=} B \times \langle w = w_0 \rangle$.



Example

Programs $prog_1$ and $prog_2$

Consider $prog_1$ and $prog_2$ as follows:

$$prog_1 \hat{=} c : \left\{ \frac{1}{2}, \langle c = H \rangle \right\}, \quad prog_2 \hat{=} c : = H \frac{1}{2} \oplus c : = T.$$

$prog_1 \sqsubseteq prog_2?$

$$\begin{aligned} & [c_0 := c] \left[c : = H \frac{1}{2} \oplus c : = T \right] \langle c = H \rangle \\ \equiv & \frac{1}{2} \times [c : = H] \langle c = H \rangle \\ & + \left(1 - \frac{1}{2} \right) \times [c : = T] \langle c = H \rangle \\ \equiv & \frac{1}{2} \times \langle H = H \rangle + \frac{1}{2} \times \langle T = H \rangle \\ \equiv & \frac{1}{2}. \end{aligned}$$



Example

Programs $prog_1$ and $prog_2$

Consider $prog_1$ and $prog_2$ as follows:

$$prog_1 \hat{=} c : \left\{ \frac{1}{2}, \langle c = H \rangle \right\}, \quad prog_2 \hat{=} c : = H \frac{1}{2} \oplus c : = T.$$

$prog_1 \sqsubseteq prog_2?$

$$\begin{aligned} & [c_0 := c] \left[c : = H \frac{1}{2} \oplus c : = T \right] \langle c = H \rangle \\ \equiv & \frac{1}{2} \times [c : = H] \langle c = H \rangle \\ & + \left(1 - \frac{1}{2}\right) \times [c : = T] \langle c = H \rangle \\ \equiv & \frac{1}{2} \times \langle H = H \rangle + \frac{1}{2} \times \langle T = H \rangle \\ \equiv & \frac{1}{2}. \end{aligned}$$



Outline

1 Motivation

- Extension to Probabilistic B
- Background

2 Our Results/Contribution

- Probabilistic specification substitution
- Fundamental theorem
- **Proof Obligations for Loops**
- Case Study



Standard rules

For a standard loop, such as

$\text{loop} \hat{=} \text{WHILE } G \text{ DO } S \text{ INVARIANT } I \text{ VARIANT } V \text{ END,}$

then $P \implies [\text{init}; \text{loop}]Q$ holds if the following are satisfied:

$$S1 \quad P \implies [\text{init}]I$$

$$S2 \quad G \wedge I \implies [S]I$$

$$S3 \quad \neg G \wedge I \implies Q$$

$$S4 \quad I \implies V \in \mathbb{N}$$

$$S5 \quad G \wedge I \implies [n := V][S](V < n)$$



Probabilistic rules

For a probabilistic loop, such as

$\text{loop} \hat{=} \text{WHILE } G \text{ DO } S \text{ INVARIANT } I \text{ EXPECTATION } E \text{ VARIANT } V \text{ END.}$

then $\langle P \rangle * A \Rightarrow [init; \text{loop}](\langle Q \rangle * B)$ holds if the following satisfies:

$P1$	$\langle P \rangle * A$	\Rightarrow	$[init](\langle I \rangle * E)$
$P1a$	P	\Rightarrow	$[init] I$
$P1b$	$\langle P \rangle * A$	\Rightarrow	$[init] E$
$P2$	$\langle G \wedge I \rangle * E$	\Rightarrow	$[S](\langle I \rangle * E)$
$P2a$	$G \wedge I$	\Rightarrow	$\llbracket S \rrbracket I$
$P2b$	$\langle G \wedge I \rangle * E$	\Rightarrow	$[S] E$
$P3$	$\langle \neg G \wedge I \rangle * E$	\Rightarrow	$\langle Q \rangle * B$
$P3a$	$\neg G \wedge I$	\Rightarrow	Q
$P3b$	$\langle \neg G \wedge I \rangle * E$	\Rightarrow	B
$P4$	I	\Rightarrow	$V \in \mathbb{N}$
$P5$	$G \wedge I$	\Rightarrow	$[n := V] \llbracket S \rrbracket (V < n)$

The difference with the previous work is that there's a clear separation between I and E .



Outline

1 Motivation

- Extension to Probabilistic B
- Background

2 Our Results/Contribution

- Probabilistic specification substitution
- Fundamental theorem
- Proof Obligations for Loops
- **Case Study**



Description of Min-Cut algorithm

Aims

- Probabilistic fundamental theorem in practice.
- Developing probabilistic system in layers.
- Analysing some of the unexpected and subtle issues.

Two phases

The algorithm is used to find the minimum cut for a connected indirect graph:

- A **cut** is a set of edges such that if we remove just those edges, the graph will become disconnected;
- A **minimum** cut is a cut with the least number of edges.

The algorithm contains two phases: **Contraction sequences** and **probabilistic amplification**.



Description of Min-Cut algorithm

Aims

- Probabilistic fundamental theorem in practice.
- Developing probabilistic system in layers.
- Analysing some of the unexpected and subtle issues.

Two phases

The algorithm is used to find the minimum cut for a connected indirect graph:

- A **cut** is a set of edges such that if we remove just those edges, the graph will become disconnected;
- A **minimum** cut is a cut with the least number of edges.

The algorithm contains two phases: **Contraction sequences** and **probabilistic amplification**.



Contraction sequences

Description

- In a **contraction step**, two connected nodes are chosen randomly and merge together.
- The probability that any specific minimum cut is kept is at least

$$\frac{N-2}{N},$$

where N is the number of nodes in the current graph.

- This step is repeated until there are two nodes left, the edges connecting the last two nodes will be the cut chosen.
- Overall, the probability that the last cut is minimum cut is at least

$$p(N) = \frac{N-2}{N} \times \frac{N-3}{N-1} \times \cdots \times \frac{2}{4} \times \frac{1}{2} = \frac{2}{N \times (N-1)}.$$



Formal development of contraction

Specification

$$ans \leftarrow \mathbf{contraction}(N) \hat{=} ans : \{ \langle pre1 \rangle * p(N), \langle ans \rangle \}$$

Implementation

```
ans ← contraction( N ) ≐
VAR n IN
  n := N; ans := TRUE;
  WHILE 2 < n DO ans ← merge(n, ans); n := n - 1 END
END
```

merge operation

$$ans \leftarrow \mathbf{merge}(n, a) \hat{=} n \in \mathbb{N} \wedge a \in \mathbf{BOOL} \quad | \quad ans := \mathbf{FALSE} \stackrel{\leq \frac{2}{n}}{\oplus} a$$


Formal development of contraction

Specification

$$ans \leftarrow \mathbf{contraction}(N) \hat{=} ans : \{ \langle pre1 \rangle * p(N), \langle ans \rangle \}$$

Implementation

```

ans ← contraction( N ) ≐
VAR n IN
  n := N; ans := TRUE;
  WHILE 2 < n DO ans ← merge(n, ans); n := n - 1 END
END

```

merge operation

$$ans \leftarrow \mathbf{merge}(n, a) \hat{=} n \in \mathbb{N} \wedge a \in \mathbf{BOOL} \quad | \quad ans := \mathbf{FALSE} \stackrel{\leq \frac{2}{n}}{\oplus} a$$


Formal development of contraction

Specification

$$ans \leftarrow \mathbf{contraction}(N) \hat{=} ans : \{ \langle pre1 \rangle * p(N), \langle ans \rangle \}$$

Implementation

```

ans ← contraction( N ) ≐
VAR n IN
  n := N; ans := TRUE;
  WHILE 2 < n DO ans ← merge(n, ans); n := n - 1 END
END

```

merge operation

$$ans \leftarrow \mathbf{merge}(n, a) \hat{=} n \in \mathbb{N} \wedge a \in \mathbf{BOOL} \quad | \quad ans := \mathbf{FALSE} \stackrel{\leq \frac{2}{n}}{\oplus} a$$


Proof obligations of contraction

Here is the summary of proof obligations for the implementation generated by the modified *B-Toolkit*.

Summary

Total	Auto Prove	BTool Prove
14	11	3



Probabilistic amplification

Description

- We repeat the contraction sequences to increase the chance of finding the right minimum cut.
- Assume that we do that M times, the probability of finding the right minimum cut is at least:

$$P(N, M) = 1 - (1 - p(N))^M .$$



Formal development of probabilistic amplification

Specification

$$ans \leftarrow \mathbf{minCut}(N, M) \hat{=} ans : \{ \langle pre2 \rangle * P(N, M), \langle ans \rangle \}$$

Implementation

```

ans ← minCut( N, M ) ≐
VAR m, a IN
  m := M; ans := FALSE;
  WHILE m ≠ 0 DO
    a ← contraction(N);
    ans := ans ∨ a;
    m := m - 1
  END
END

```



Formal development of probabilistic amplification

Specification

$$ans \leftarrow \mathbf{minCut}(N, M) \hat{=} ans : \{ \langle pre2 \rangle * P(N, M), \langle ans \rangle \}$$

Implementation

```

ans ← minCut( N, M ) ≐
VAR m, a IN
  m := M; ans := FALSE;
  WHILE m ≠ 0 DO
    a ← contraction(N);
    ans := ans ∨ a;
    m := m - 1
  END
END

```



Proof obligations of probabilistic amplification

Here is the summary of proof obligations for probabilistic amplification produced by the modified *B-Toolkit*.

Summary

Total	Auto Prove	BTool Prove
14	13	0

Problem?

There is one proof obligation that cannot be proved.

Solution

The problem observed is due to the fact that in the definition for probabilistic specification substitution, we **did not specify termination**.

In *B*, termination of all programs must be proved, so we should introduce **terminating probabilistic specification substitution** and its **fundamental theorem**.



Proof obligations of probabilistic amplification

Here is the summary of proof obligations for probabilistic amplification produced by the modified *B-Toolkit*.

Summary

Total	Auto Prove	BTool Prove
14	13	0

Problem?

There is one proof obligation that cannot be proved.

Solution

The problem observed is due to the fact that in the definition for probabilistic specification substitution, we **did not specify termination**.

In *B*, termination of all programs must be proved, so we should introduce **terminating probabilistic specification substitution** and its **fundamental theorem**.



Proof obligations of probabilistic amplification

Here is the summary of proof obligations for probabilistic amplification produced by the modified *B-Toolkit*.

Summary

Total	Auto Prove	BTool Prove
14	13	0

Problem?

There is one proof obligation that cannot be proved.

Solution

The problem observed is due to the fact that in the definition for probabilistic specification substitution, we **did not specify termination**.

In *B*, termination of all programs must be proved, so we should introduce **terminating probabilistic specification substitution** and its **fundamental theorem**.



Summary

- **Abstractly specify and refine** probabilistic system.
- Development can be separated into **layers**.
- **Termination** condition is checked when developing systems using the *B-Toolkit*.

- Future work
 - Multiple expectations.
 - Fundamental theorem for refining system with multiple expectations.



For Further Reading I



C. Morgan and A. McIver.

Abstraction, Refinement and Proof for Probabilistic Systems.
Springer-Verlag, 2004.



T.S. Hoang, Z. Jin, K. Robinson, C. Morgan and A. McIver.
Probabilistic Invariant for Probabilistic Machines.

Proceedings of the 3rd International Conference of B and Z Users, volume 2651 of *LNCS*, 2003.



N. White.

Probabilistic Specification and Refinement
Master Thesis, Keble College, 1996.



M.S. Ying.

Reasoning about probabilistic sequential programs in a
probabilistic logic.

Acta Informatica, volume 39, 2003.

