CALIFORNIA INSTITUTE OF TECHNOLOGY

EARTHQUAKE ENGINEERING RESEARCH LABORATORY

UNCERTAINTY PROPAGATION AND FEATURE SELECTION FOR LOSS ESTIMATION IN PERFORMANCE-BASED EARTHQUAKE ENGINEERING

BY

JIANYE CHING, KEITH A. PORTER & JAMES L. BECK

REPORT NO. EERL 2004-02

PASADENA, CALIFORNIA

FEBRUARY 2004

Acknowledgements

The authors gratefully acknowledge the financial support of Caltech's George W. Housner Postdoctoral Fellowship (J. Ching), the George W. Housner Senior Research Fellowship (K. Porter), and the Pacific Earthquake Engineering Research Center under Award Number EEC-9701568 of the Earthquake Engineering Research Centers Program of the National Science Foundation.

ABSTRACT

This report presents a new methodology, called moment matching, of propagating the uncertainties in estimating repair costs of a building due to future earthquake excitation, which is required, for example, when assessing a design in performance-based earthquake engineering. Besides excitation uncertainties, other uncertain model variables are considered, including uncertainties in the structural model parameters and in the capacity and repair costs of structural and non-structural components. Using the first few moments of these uncertain variables, moment matching requires only a few well-chosen point estimates to propagate the uncertainties to estimate the first few moments of the repair costs with high accuracy. Furthermore, the use of moment matching to estimate the exceedance probability of the repair costs is also addressed. These examples illustrate that the moment-matching approach is quite general; for example, it can be applied to any decision variable in performance-based earthquake engineering.

Two buildings are chosen as illustrative examples to demonstrate the use of moment matching, a hypothetical three-story shear building and a real seven-story hotel building. For these two examples, the assembly-based vulnerability approach is employed when calculating repair costs. It is shown that the moment-matching technique is much more accurate than the well-known First-Order-Second-Moment approach when propagating the first two moments, while the resulting computational cost is of the same order. The repair-cost moments and exceedance probability estimated by the moment-matching technique are also compared with those by Monte Carlo simulation. It is concluded that as long as the order of the moment matching is sufficient, the comparison is satisfactory. Furthermore, the amount of computation for moment matching scales only linearly with the number of uncertain input variables.

Last but not least, a procedure for feature selection is presented and illustrated for the second example. The conclusion is that the most important uncertain input variables among the many influencing the uncertainty in future repair costs are, in order of importance, ground-motion spectral acceleration, component capacity, ground-motion details and unit repair costs.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 METHODS TO PROPAGATE UNCERTAINTY IN EARTHQUAKE REPAIR COST MODELS

Implementation of performance-based earthquake engineering (PBEE) often requires the quantification of uncertain future repair costs associated with earthquake damage (e.g. Porter 2003). Furthermore, real-estate investment and business recovery decision-making could also benefit from reliable techniques to estimate future repair costs (Beck et al, 2002, 1999; Porter et al, 2004). These future repair costs are uncertain because they depend on uncertain basic variables such as the occurrence dates and times; shaking intensities, and other aspects of the earthquake ground motions at a site; the dynamic properties of the facility; and the capacities and repair costs of facility components. Each of these uncertainties influence the subsequent ones, as they propagate from earthquake occurrence, to ground motion at a site, to structural response, to damage, and finally to repair cost. A reliable repair-cost estimate based on a model of this entire process must therefore account for this propagation of uncertainties.

Four methods have been used to propagate uncertainties for repair costs: deterministic sensitivity studies; Monte Carlo simulation (MCS); first-order, second-moment (FOSM); and FORM or SORM (first-order or second-order reliability method):

(a) Deterministic sensitivity studies can be used to identify those input variables that most strongly affect uncertainty in an output parameter of interest. Porter et al. (2002a and 2002b) apply this method to examine the contributions of each uncertain input variable to the uncertainty in repair costs by using a graphic depiction called a tornado diagram.

(b) MCS techniques are simple, effective means to propagate uncertainties and to explore the probability distribution of an output parameter. They can be computationally expensive for very large systems or for situations where one wishes to explore low-probability events. Porter et al. (2002c) and Beck et al. (2002) present two recent examples of Monte Carlo simulation for earthquake loss to individual buildings.

(c) FOSM (Melchers 1999) is a convenient and efficient technique for propagating uncertainty where one is primarily interested in the mean and variance of an output variable rather than the tails of the distribution. Baker and Cornell (2003) summarize such a technique for estimating future earthquake losses. FOSM is less computationally expensive than MCS, but as its name implies, can achieve only first-order accuracy and it is not clear how to improve its accuracy. It cannot handle higher moments and requires quantifications of the correlations of intermediate uncertain variables.

(d) First-order and second-order reliability methods (FORM/SORM) are powerful techniques for estimating the probability of rare events, i.e., for exploring the tails of a loss distribution. Like FOSM, FORM/SORM also requires an estimate of correlation between intermediate variables. Haukaas and Der Kiureghian (2003) summarize a FORM approach to estimating failure probability in a finite-element reliability problem. To our knowledge, FORM and SORM have not yet been used to estimate failure probability for a structural system where the performance function is expressed in terms of repair costs exceeding a prescribed value.

1.2 NEEDED: EFFICIENT UNCERTAINTY PROPAGATION IN A HIGH-ACCURACY LOSS MODEL

Missing from these options is a computationally efficient technique to propagate uncertainty that does not require an estimate of correlation between intermediate variables and that can provide arbitrarily high order moments of repair costs. In this report, we re-cast the uncertainty propagation problem into a general analysis problem. Using a Taylor-series expansion, we propose a moment-matching technique that propagates uncertainties accurately with a few carefully selected sample points in the input-variable space. We show that the new method possesses all of the afore-mentioned advantages of FOSM and MCS while avoiding many of their limitations. With the moment-matching technique, we study four types of quantitative information regarding repair costs in this report:

1. The central moments (e.g., the mean and variance) of repair costs during the next T years.
2. The mean exceedance frequency of single-event repair costs over some threshold during the next T years.
3. The central moments of the repair costs due to the extreme event in the next T years.
4. The exceedance probability of repair costs due to the extreme event over some threshold in the next T years.

Feature selection is also examined, i.e., determining the degree of importance of the uncertainty in each input variable on the repair cost estimate. The benefit of doing this is twofold: first, the input variables with little importance can be considered deterministic to reduce calculation efforts. Second, the input variables that play major roles can be further studied to improve our knowledge of them.

The problem of feature selection has been studied by Porter et al. (2002a b) using a deterministic sensitivity analysis approach, in which an input variable for the loss modeling is considered to be important if and only if repair costs are sensitive to its change. In this report, we reconsider the feature selection problem under a probabilistic approach built on the moment-matching technique. We utilize a measure of information change before and after the removal of the uncertainties of an input variable. An input variable is considered to be important if and only if the corresponding information change is large. This measure can potentially detect any change in the probability density function (PDF) of repair costs before and after the removal of the uncertainty of an input variable.

1.3 ORGANIZATION OF THE REPORT

The structure of this report is as follows: Section 2 presents an overview of performance-based earthquake engineering (PBEE) as it is currently formulated by the Pacific Earthquake Engineering Research (PEER) Center. Section 2 also defines the problem of uncertainty propagation under this PBEE framework. In Section 3, we briefly discuss the FOSM technique of propagating uncertainty and describe in detail the new moment-matching technique. The performance of the new technique and other techniques is compared using simulations. In Section 4, we derive algorithms for calculating the moments and exceedance probability of repair costs due to a future earthquake. Based on these results, we present procedures for estimating the four types of quantitative information listed in Section 1.2. In Section 5, we present the principle and formulas for feature selection.

We illustrate the moment-matching technique with two examples of uncertainty propagation under the PEER PBEE framework. The first example (Section 4.5) is about estimating the repair cost measures discussed above for a simple three-story shear building. The main focus of this example is to compare the performances of the new moment-matching technique and other techniques.

The second example (Chapter 6) illustrates the estimation of moments and exceedance probability of the single-event repair costs of a non-ductile reinforced concrete moment-frame building (a 1960s-era hotel building in Van Nuys, California) due to the extreme earthquake event in the next 50 years. The main focus of this example is to demonstrate the new moment-matching technique on a real building and to illustrate the feature selection procedure.

2 PERFORMANCE-BASED EARTHQUAKE ENGINEERING

Estimating economic losses of a structure due to future earthquakes is a major part of performance-based earthquake engineering (PBEE). The Pacific Earthquake Engineering Research (PEER) Center has proposed the framework shown in Figure 2-1 (Porter, 2003) for its second-generation PBEE methodology.

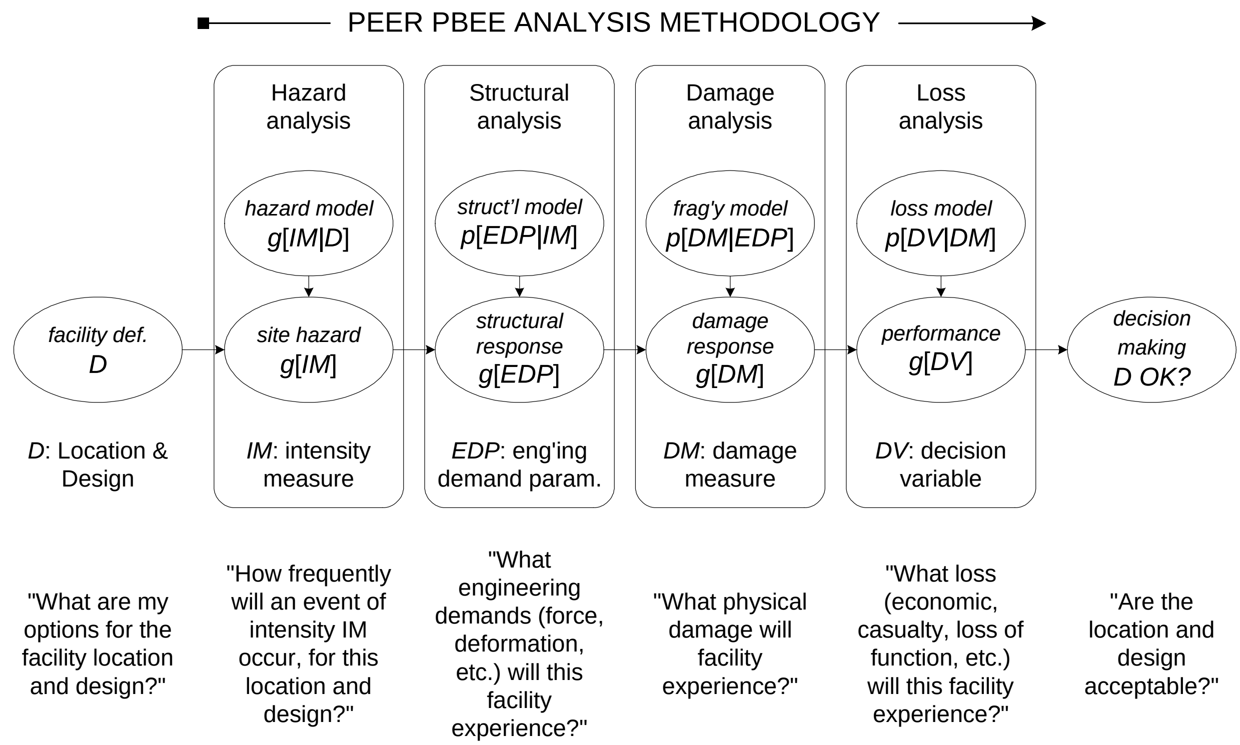

Figure 2-1. PEER's PBEE framework

As discussed in Porter (2003), PEER's PBEE approach begins with a definition of the facility to be analyzed (denoted here by D), and involves four analytical stages: hazard analysis, structural analysis, damage analysis, and loss analysis. In the figure, the expression $p[X|Y]$ refers to the probability density of X conditioned on knowledge Y, and $g[X|Y]$ refers to the mean occurrence frequency of X given Y (equivalent to the absolute value of the first derivative of the frequency with which X is exceeded, given Y). Equation (2.1) frames the PEER methodology mathematically. Note that Figure 2-1 omits conditioning on *D* after the hazard analysis for brevity, but it is nonetheless implicit.

$$g[DV] = \iiint p[DV \mid DM, D] p[DM \mid EDP, D] p[EDP \mid IM, D] g[IM \mid D] d(IM) d(EDP) d(DM) \qquad (2.1)$$

Facility definition. To define the facility one must know its location (latitude and longitude) and design, including site soils, substructure, structural and nonstructural components, jointly denoted by *D*. One creates an inventory of the damageable assemblies and identifies the engineering demand parameter — *EDP*, which might be story drift ratio, member force, etc. — that would cause damage to each assembly.

Hazard analysis. In the hazard analysis, one considers the seismic environment (nearby faults, their magnitude-frequency recurrence rates, mechanism, site distance, site conditions, etc.) and evaluates the seismic hazard at the facility considering *D*, to produce the seismic hazard, $g[IM|D]$, where *IM* refers to the intensity measure. *IM* can be parameterized in any of a variety of terms, such as peak horizontal ground acceleration, Arias intensity, etc. It is common to use $S_a(T_1)$, the damped elastic spectral acceleration at the small-amplitude fundamental period of the structure, which is readily available by using software such as Frankel and Leyedecker (2001), adjusting to account for site classification such as by using F_a or F_v as appropriate from the 2000 International Building Code (International Code Council, 2000). In the present analysis, we use $S_a(T_1)$ for *IM*.

Structural analysis. In the structural analysis, the engineer creates a structural model of the facility in order to estimate the uncertain structural response, measured in terms of a vector of engineering demand parameters (*EDP*), conditioned on seismic excitation and design ($p[EDP|IM,D]$). *EDP*s can include internal member forces or local or global deformations, including ground failure (a partial list of *EDP*s in use by PEER is provided in Porter, 2002). The structural analysis typically takes the form of a series of nonlinear time-history structural analyses using a suite of strong-motion records that are scaled to have the specified *IM*. The structural model need not be deterministic — some PEER analyses have included uncertainty in the mass, damping, and force-deformation characteristics of the model. The present study does so, as will be discussed later.

Damage analysis. *EDP* is then input to a set of fragility functions that model the probability of various levels of physical damage (expressed via damage measures, or *DM*), conditioned on structural response and design, $p[DM|EDP,D]$. Physical damage is not described at a detailed level, but instead is defined relative to particular repair efforts required to restore the component to its undamaged state. Fragility functions currently in use give the probability of various levels of damage to individual beams, columns, nonstructural partitions, or pieces of laboratory equipment, as functions of various internal member forces, story drift, etc. These functions are drawn from laboratory or field experience. For example, PEER has compiled a library of destructive tests of reinforced concrete columns (Eberhard et al., 2001). The result of the damage analysis is a probabilistic vector of *DM*. Note that component damage may be correlated with structural characteristics of *D*, even conditioned on *EDP*.

Loss analysis. The last stage in the analysis is the probabilistic estimation of performance (parameterized via various decision variables, *DV*), conditioned on damage and design $p[DV|DM,D]$. Decision variables measure the seismic performance of the facility in terms of greatest interest to stakeholders, whether in dollars, deaths, downtime, or other metrics. Dollar losses can be estimated using standard construction-contracting principles, given the detailed damage state of the facility. Deaths can be estimated using empirical casualty estimates, as discussed by Seligson and Shoaf (2002). Repair duration can be estimated using construction scheduling principles, as discussed in Porter (2000).

Decision-making. The analysis produces estimates of the frequency with which various levels of *DV* are experienced, given the facility definition *D*. These frequencies can be used to inform a variety of risk-management decisions. For example, a common concern among insurers the need for reinsurance to deal with catastrophically high losses. Consequently, it is of interest to know the frequency with which future repair cost will exceed some ruin threshold, $G[DV|R]$, where $G[X|Y]$ refers to the frequency with which X is exceeded, conditioned on knowledge Y. For an individual facility exposed to seismic risk, one can calculate this ruin frequency as

$$G\left[R \mid D\right] = \int_{DV=R}^{\infty} g\left[DV \mid D\right] dDV \tag{2.2}$$

Defining the problem of uncertainty propagation. Observe that *DV* can be viewed as a deterministic function of a number *n* of uncertain input variables. For example, if we are concerned with uncertain future repair costs, we can consider the hazard model as an uncertain parameter *IM*, the structural model as a as a set of one or more structural variables (*SM*), the fragility model as a set of uncertain capacities *C*, and the loss model as a set of unit repair costs (*URC*), etc, collectively denoted by *X*, i.e.,

$$DV = f(X) \tag{2.3}$$

where $X = \{IM, SM, C, URC, \ldots\} \in R^n$ contains all basic variables. The problem of uncertainty propagation under PEER's PBEE framework can be defined as follows: Given the moments or joint PDF of *X* (we will call it the *X* PDF), the goal is to determine the moments of *DV* or the probability that *DV* will exceed a threshold value (i.e. determine the PDF of *DV*).

3 UNCERTAINTY PROPAGATION

We first briefly discuss the FOSM technique of propagating uncertainties, then present the moment-matching (MM) technique. We show theoretically that the latter is more accurate than the former. The presentation of the two techniques is followed by several simple examples to demonstrate the effectiveness of the MM technique.

3.1 FIRST-ORDER SECOND-MOMENT TECHNIQUE

The FOSM technique assumes the function $f(X)$ in Equation (2.3) for the decision variable DV is roughly linear in $X \in \mathbb{R}^n$ in the support region of the X PDF so that the first two central moments of DV, $E(DV)$ and $Var(DV)$, are simple functions of these moments for X. For the FOSM technique, it is assumed that the Taylor series expansion of $f(X)$ around $X = EX$ exists, then:

$$DV = f(X) = f(EX) + D_x f + D_x^2 f/2! + D_x^3 f/3! + D_x^4 f/4! \ldots, \tag{3.1}$$

where

$$D_x^i f \equiv \left(\sum_{j=1}^{n} (X_j - EX_j) \cdot (\partial/\partial x_j) \right)^i f(x) \bigg|_{x=EX} . \tag{3.2}$$

To verify the above multidimensional Taylor series expansion, consider expanding $f(X)$ in the s direction, where $s = (X - EX)/\|X - EX\|$ is the unit vector in the (X-EX)-direction. With the one-dimensional Taylor series expansion, we have

$$f(X) = f(EX) + \sum_{i=1}^{\infty} \frac{\partial^i f(x)}{\partial s^i} \bigg|_{x=EX} \cdot \frac{\|X - EX\|^i}{i!} , \tag{3.3}$$

where the term $\partial^i f(x)/\partial s^i \big|_{X=EX}$ is equal to

$$\left.\frac{\partial^i f(x)}{\partial s^i}\right|_{X=EX} = \left(s\cdot\nabla\right)^i f(x)\Big|_{x=EX}$$

$$= \left(s_1\frac{\partial}{\partial x_1}+\ldots+s_n\frac{\partial}{\partial x_n}\right)^i f(x)\Bigg|_{x=EX} = \left(\sum_{j=1}^{n}\frac{X_j - EX_j}{\|X-EX\|}\cdot(\partial/\partial x_j)\right)^i f(x)\Bigg|_{x=EX} \quad . \tag{3.4}$$

Substitute Equation (3.4) into Equation (3.3), then Equation (3.2) follows.

As a result, the first two moments of *DV* are

$$E(DV) = \underbrace{f(EX)}_{0th} + \underbrace{E(D_x f)}_{1st} + \underbrace{E(D_x^2 f/2!)}_{2nd} + \underbrace{E(D_x^3 f/3!)}_{3rd} + \underbrace{E(D_x^4 f/4!)}_{4th}\ldots$$

and (3.5)

$$Var(DV) = E\left([DV - E(DV)]^2\right)$$
$$= \underbrace{E\left((D_x f)^2\right)}_{2nd} + \underbrace{2E\left((D_x f)(D_x^2 f/2!)\right)}_{3rd} + \underbrace{2E\left((D_x f)(D_x^3 f/3!)\right) + E\left((D_x^2 f/2!)(D_x^2 f/2!)\right) - \left[E(D_x^2 f/2!)\right]^2}_{4th} + \ldots$$

Under the assumption that $f(X)$ is linear in the support region of the X PDF, all second or higher derivatives of $f(X)$ with respect to X vanish; therefore, $E(DV)$ and $Var(DV)$ are approximated by:

$$E(DV)_{FOSM} = \underbrace{f(EX)}_{0th} + \underbrace{E(D_x f)}_{1st} = f(EX)$$

and (3.6)

$$Var(DV)_{FOSM} = \underbrace{E\left((D_x f)^2\right)}_{2nd} = (\nabla_x f|_{x=EX})\cdot Var(X)\cdot(\nabla_x f|_{x=EX})^T$$

where $E(DV)_{FOSM}$ and $Var(DV)_{FOSM}$ denote the FOSM estimates for $E(DV)$ and $Var(DV)$, and $\nabla_x f \in R^{1\times n}$ is the Jacobian matrix.

The approximations $E(DV)_{FOSM}$ and $Var(DV)_{FOSM}$ are accurate estimates of $E(DV)$ and $Var(DV)$ if $f(X)$ is almost linear on the support region of the X PDF, and the FOSM approximations become exact when $f(X)$ is indeed linear in X. We say that the accuracy of the FOSM technique is first-order since the technique is exact if $f(X)$ is a first-order polynomial. On the other hand, the approximations are poor if $f(X)$ is highly nonlinear on the support region of X PDF.

3.2 MOMENT-MATCHING TECHNIQUE

An alternative method for uncertainty propagation is the moment-matching (MM) technique. It is a point-estimate method first proposed by Rosenblueth (1975). Also see Zhao and Ono (2000) and Julier et al. (2000) for recent developments. The MM technique is a procedure where the PDF of X is modeled by an "equivalent" discrete PDF containing several weighted delta functions, where the one-dimensional delta function δ(.) is a generalized (or symbolic) function that has the properties:

$$\begin{aligned} &\delta(x-a)=0 \qquad x \neq a \\ &\int_{-\infty}^{\infty} \delta(x)\,dx = 1 \end{aligned} \tag{3.7}$$

i.e., it imparts a unit "impulse" to the system at $x = a$, but is zero for all other values of x. The delta functions of the equivalent discrete PDF have specific positions and weights such that the first few moments of this PDF match those of X. Thus, rather than approximating the function $f(x)$ in Equation (2.3) by some simplified functional form, as in the FOSM technique, the MM technique approximates the X PDF with a discrete PDF by matching specified moments of X. Moreover, we show that the MM technique has the potential to propagate uncertainties more accurately than the FOSM technique. As a simple illustration, one could replace a standard Gaussian distribution of Figure 3-1(a) with the PDF of three weighted delta functions whose positions and weights are shown in Figure 3-1(b). The first three moments of the two distributions are equal, although their higher moments are not.

Consider $Y = g(X)$ and we would like to estimate *E*(*Y*). According to Equation (3.5), we have

$$E(Y) = \underbrace{g(EX)}_{0th} + \underbrace{E(D_x g)}_{1st} + \underbrace{E(D_x^2 g/2!)}_{2nd} + \underbrace{E(D_x^3 g/3!)}_{3rd} + \underbrace{E(D_x^4 g/4!)}_{4th} \ldots \tag{3.8}$$

Note that if we can estimate *E*(*Y*) for general *g*(*X*), we can estimate all moments of *DV*=*f*(*X*) by letting $g(X) = f(X)^r$ for the r^{th} moment of *DV*.

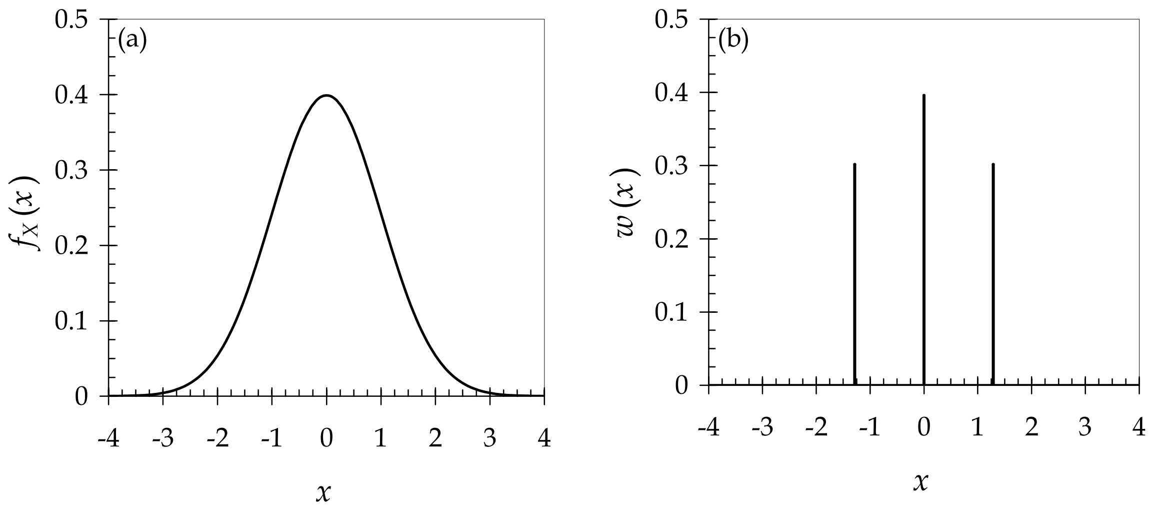

Figure 3-1. Simple illustration of moment-matching approach

The key observation for the MM technique is as follows: the i^{th} order Taylor series term in Equation (3.8), which is

$$E(D_x^i g) \equiv E\left(\sum_{j=1}^{n} (X_j - EX_j)\cdot(\partial/\partial x_j)\right)^i g(x)\Bigg|_{x=EX}, \tag{3.9}$$

depends on the i^{th} central moments of the *X* PDF and the i^{th}-order derivatives of *g*(*X*). In the MM framework, a pseudo-PDF consisting of weighted delta functions is used to match the first *p* central moments of the *X* PDF. (The prefix “pseudo” is used because in some situations we allow the weights of the delta functions to be negative). Since this pseudo-PDF only consists of weighted delta functions, the propagation of this pseudo-PDF through $g(X)$ can be done analytically, and the corresponding pseudo-PDF of Y also contains weighted delta functions. The mean of this pseudo-PDF of Y, called the MM estimate of $E(Y)$, can be again computed easily and analytically, since the pseudo-PDF contains only weighted delta functions.

In the MM procedure we use the exact functional form for $g(X)$ rather than a linear (or higher order) approximation. Therefore, if we can match up to the p^{th} moment of the X PDF, we are able to characterize up to the p^{th}-order Taylor series term of $E(Y)$. In what follows, we present the procedures of the MM technique for the special case that X is one-dimensional and discuss the idea of the technique. The extension for multi-dimensional X will be described later.

3.2.1 Moment Matching: One-dimensional X

Step 1: Characterize the first p moments of the X PDF: Suppose that we know the first p moments of the X PDF. It is straightforward to find a pseudo PDF that contains q weighted delta functions and has identical first p moments as the X PDF. Let X' denote the uncertain variable associated with the delta-function PDF:

$$p_{X'}(x) = \sum_{i=1}^{q} w_i \delta(x - \chi_i) \qquad \sum_{i=1}^{q} w_i = 1 \tag{3.10}$$

By careful selection of the number q of delta functions, their weights $\{w_i : i = 1,...q\}$ and locations $\{\chi_i : i = 1,...q\}$, we can assure that the delta-function PDF has the first p central moments identical to those of the X PDF. To do this, we need to solve the following nonlinear equations for the weights and locations:

$$\begin{aligned} &\sum_{i=1}^{q} w_i \left(\chi_i - \left(\sum_{j=1}^{q} w_j \chi_j \right) \right)^k = E\left((X - EX)^k \right) \qquad k = 1, 2, \ldots p \\ &\sum_{i=1}^{q} w_i = 1 \end{aligned} \tag{3.11}$$