1.4 : Expectation and Other Measures of a Distribution

1.4.1 : Expected value

The **expectation** or **expected value** of a continuous RV, X, is defined as its integral over the pdf of X. It is usually denoted by $E(X)$. Thus

$$E(X) = \int_{-\infty}^{\infty} x\, p(x) dx$$

Similarly, the expected value of a discrete (e.g. Poisson) RV is defined by

$$E(X) = \sum_{x=0}^{\infty} x\, p(x)$$

The expected value is also known as the **mean**, and is often written as $\overline{x}$, or $<x>$.

1.4.2 : Median value

The **median** of a RV, X, is the value, $x_{\rm med}$, which divides the CDF into two equal halves. Thus $x_{\rm med}$ satisfies

$$\int_{-\infty}^{x_{\rm med}} p(x) dx = 0.5$$

If the PDF is symmetric about the mean, then the mean and median are identical.

1.4.3 : Modal value

The **mode** of a RV, X, is the value of X at the maximum of the PDF. Thus $x_{\rm mode}$ satisfies

$$\frac{\partial p(x = x_{\rm mode})}{\partial x} = 0$$

Obviously the mode may not be uniquely defined. For example, for $U(a,b)$, $\partial p / \partial x = 0$ for all $x \in (a,b)$.

1.4.4 : Variance

The **variance** of X is defined as (for a continuous RV)

$$\text{var}(X) = \int_{-\infty}^{\infty} (x - \overline{x})^2\, p(x) dx$$

with the analogous expression for a discrete RV. The variance is usually denoted by σ^2, while $\sigma = \sqrt{(\sigma^2)}$ is called the **standard deviation**.

For either continuous or discrete RVs the following equation holds

$$\mathrm{var}(X) \quad = \quad E(X^2) - [E(X)]^2$$

(The proof of this result is left as an exercise).

1.4.5 : Examples

The following table summarises the mean value and the variance of the uniform and normal distribution. (Proofs are left as an exercise; all results are quite straightforward to derive).

X	$p(x)$	$E(X)$	$\mathrm{var}(X)$
Poisson	$\frac{(\mu)^x}{x!}\, e^{-\mu}$	μ	μ
Uniform	$1/(b-a)$	$(a+b)/2$	$(b-a)^2/12$
Normal	$\frac{1}{\sqrt{2\pi}\sigma}\, \exp[-\frac{1}{2\sigma^2}(x-\mu)^2]$	μ	σ^2

The next two measures of a distribution are particular expectation values.

1.4.6 : Skewness and Kurtosis

The (normalised) **skewness** of X, is defined by

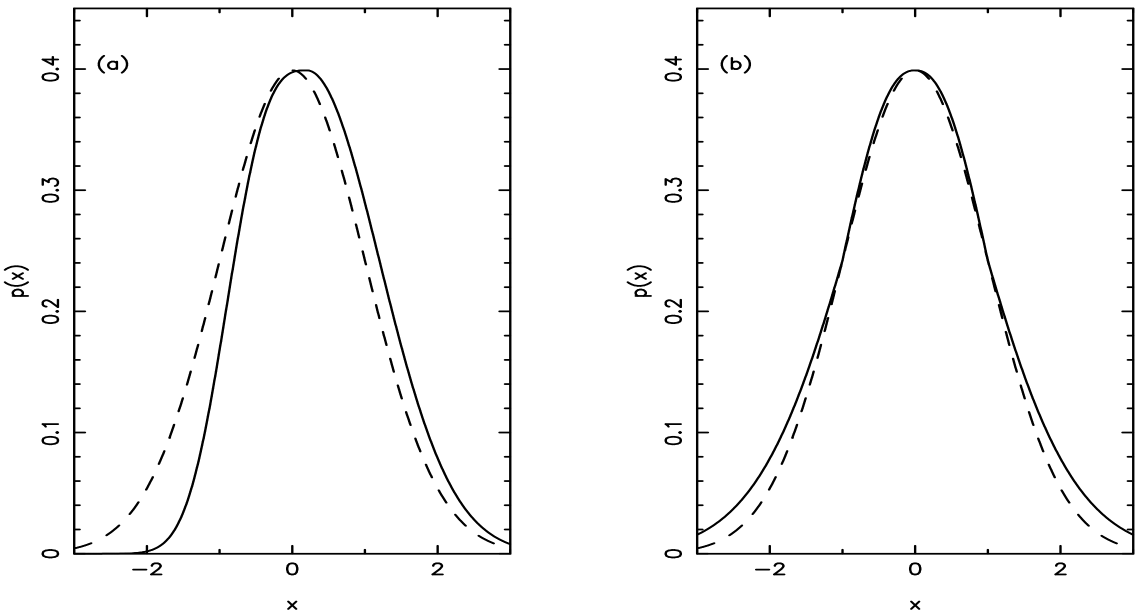
$$\mathrm{skew}(X) \quad = \quad E[(X-\overline{x})^3]/[\mathrm{var}(X)]^{\frac{3}{2}}$$

In a similar manner, the (normalised) **kurtosis** of X is defined by

$$\mathrm{kurt}(X) \quad = \quad E[(X-\overline{x})^4]/[\mathrm{var}(X)]^2 \quad - \quad 3$$

For a normally distributed RV skew(X) and kurt(X) are identically zero. The measured skewness and kurtosis of a sample of real data is often used as a test of whether those data are *drawn* from a normal distribution (see Section 2). If skew(X) > 0 then the PDF of X is 'positively' lopsided. If kurt(X) > 0 then the PDF of X has wider tails than a normally distributed RV (see Figure 6).

Figure 6: Examples of PDFs with positive skewness (a) and kurtosis (b). PDFs are shown as solid curves; Gaussian distributions are shown as dashed curves for comparison.



1.4.7 : Variance of a Function of a RV

The variance, var$[f(X)]$, of an arbitrary function of X can be approximated to second order by the following expression

$$\text{var}[f(X)] \quad = \quad \text{var}(X) \left(\frac{\partial f}{\partial x}\right)^2_{x=\overline{x}}$$

This expression is often used to assign an error to a function of a random variable. For example, suppose an experiment involves measuring a RV, X, but one wishes to determine the variance of $Y = X^2$. The above formula tells us that

$$\text{var}[Y] \quad = \quad \text{var}(X)\,[2\overline{x}]^2 \quad = \quad 4\text{var}(X)\overline{x}^2$$

(For a proof of this result, see handout on webpage). To determine the *distribution* of Y we need to define a **variable transformation**.