2.8 EXPECTATION AND VARIANCE

For a **discrete** random variable X taking values in set $\mathbb{X}$ with mass function f_X, the **expectation** of X is defined by

$$\mathrm{E}_{f_X}[\,X\,] = \sum_{x \in \mathbb{X}} x f_X(x) \equiv \sum_{x=-\infty}^{\infty} x f_X(x)$$

as $f_X(x) \equiv 0$ for $x \notin \mathbb{X}$.

For a **continuous** random variable X taking values in interval $\mathbb{X}$ with pdf f_X, the expectation of X is defined by

$$\mathrm{E}_{f_X}[X] = \int_{\mathbb{X}} x f_X(x)\, dx \equiv \int_{-\infty}^{\infty} x f_X(x)\, dx$$

as $f_X(x) \equiv 0$ for $x \notin \mathbb{X}$.

The **variance** of X is defined by

$$\mathrm{Var}_{f_X}[X] = \mathrm{E}_{f_X}[(X - \mathrm{E}_{f_X}[X])^2] = \mathrm{E}_{f_X}\left[X^2\right] - \{\mathrm{E}_{f_X}[X]\}^2 .$$

Interpretation : The expectation and variance of a probability distribution can be used to aid description, or to characterize the distribution; the EXPECTATION is a measure of **location** (that is, the "centre of mass" of the probability distribution. The VARIANCE is a measure of **scale** or **spread** of the distribution (how widely the probability is distributed) .

NOTES
(i) Take care when carrying out summation/integration to only include non-zero terms/integrand.
(ii) Sum/integral calculations may be **infinite**.

EXAMPLE Suppose that X is a discrete Poisson random variable taking values on $\mathbb{X} = \{0, 1, 2, ...\}$ with pdf

$$f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda} \qquad x = 0, 1, 2, ...$$

and zero otherwise.

Then

$$\mathrm{E}_{f_X}[\,X\,] \;=\; \sum_{x=-\infty}^{\infty} x f_X(x) = \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

using the power series expansion definition for the exponential function

$$e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + ...$$

EXAMPLE Suppose that X is a continuous random variable taking values on $\mathbb{X} = \mathbb{R}^+$ with pdf

$$f_X(x) = \frac{2}{(1+x)^3} \qquad x > 0.$$

Then, integrating by parts.

$$\mathrm{E}_{f_X}[\,X\,] = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_0^{\infty} \frac{2x}{(1+x)^3}\, dx = \left[-\frac{x}{(1+x)^2}\right]_0^{\infty} + \int_0^{\infty} \frac{1}{(1+x)^2}\, dx = 0 - \left[-\frac{1}{1+x}\right]_0^{\infty} = 1$$

2.8.1 RESULTS INVOLVING EXPECTATIONS: SUMS OF RANDOM VARIABLES:

Suppose that X_1 and X_2 are independent random variables, and a_1 and a_2 are constants. Then if $Y = a_1X_1+a_2X_2$, it can be shown that

$$\begin{aligned}
\mathrm{E}_{f_Y}[Y] &= a_1\mathrm{E}_{f_{X_1}}[X_1] + a_2\mathrm{E}_{f_{X_2}}[X_2] \\
\mathrm{Var}_{f_Y}[Y] &= a_1^2\mathrm{Var}_{f_{X_1}}[X_1] + a_2^2\mathrm{Var}_{f_{X_2}}[X_2]
\end{aligned}$$

so that, in particular (when $a_1 = a_2 = 1$) we have

$$\begin{aligned}
\mathrm{E}_{f_Y}[Y] &= \mathrm{E}_{f_{X_1}}[X_1] + \mathrm{E}_{f_{X_2}}[X_2] \\
\mathrm{Var}_{f_Y}[Y] &= \mathrm{Var}_{f_{X_1}}[X_1] + \mathrm{Var}_{f_{X_2}}[X_2]
\end{aligned}$$

so we have a simple additive property for expectations and variances. Note also that if $a_1 = 1, a_2 = -1$, then

$$\begin{aligned}
\mathrm{E}_{f_Y}[Y] &= \mathrm{E}_{f_{X_1}}[X_1] - \mathrm{E}_{f_{X_2}}[X_2] \\
\mathrm{Var}_{f_Y}[Y] &= \mathrm{Var}_{f_{X_1}}[X_1] + \mathrm{Var}_{f_{X_2}}[X_2]
\end{aligned}$$

Sums of random variables crop up naturally in many statistical calculations. Often we are interested in a random variable Y that is defined as the sum of some other **independent and identically distributed** (i.i.d) random variables, $X_1, ..., X_n$. If

$$Y = \sum_{i-1}^{n} X_i \qquad \text{with} \qquad \mathrm{E}_{f_{X_i}}[X_i] = \mu \quad \text{and} \quad \mathrm{Var}_{f_{X_i}}[X_i] = \sigma^2$$

we have

$$\mathrm{E}_{f_Y}[Y] = \sum_{i-1}^{n} \mathrm{E}_{f_{X_i}}[X_i] = \sum_{i-1}^{n} \mu = n\mu \qquad \text{and} \qquad \mathrm{Var}_{f_Y}[Y] = \sum_{i-1}^{n} \mathrm{Var}_{f_{X_i}}[X_i] = \sum_{i-1}^{n} \sigma^2 = n\sigma^2$$

and also, if

$$\overline{X} = \frac{1}{n}\sum_{i-1}^{n} X_i \qquad \text{is the } \textbf{sample mean} \text{ random variable}$$

then, using the properties listed above

$$\mathrm{E}_{f_{\overline{X}}}[\overline{X}] = \frac{1}{n}\mathrm{E}_{f_Y}[Y] = \frac{1}{n}n\mu = \mu \qquad \text{and} \qquad \mathrm{Var}_{f_Y}[Y] = \frac{1}{n^2}\mathrm{Var}_{f_Y}[Y] = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

EXPECTATIONS OF A FUNCTION OF A RANDOM VARIABLE

Suppose that X is a random variable, and $g(.)$ is some function. Then we can define the expectation of $g(X)$ (that is, the expectation of a function of a random variable) by

$$\mathrm{E}_{f_X}[\,g(X)\,] = \begin{cases} \displaystyle\sum_{x=-\infty}^{\infty} g(x)f_X(x) & \text{DISCRETE CASE} \\ \displaystyle\int_{-\infty}^{\infty} g(x)f_X(x)\,dx & \text{CONTINUOUS CASE} \end{cases}$$

For example, if X is a continuous random variable, and $g(x) = \exp\{-x\}$ then

$$\mathrm{E}_{f_X}[\,g(X)\,] = \mathrm{E}_{f_X}[\,\exp\{-X\}\,] = \int_{-\infty}^{\infty} \exp\{-x\}\, f_X(x)\,dx$$

Note that $Y = g(X)$ is also a random variable whose probability distribution we can calculate from the probability distribution of X.

2.8.2 EXPECTATION AND VARIANCE RESULTS FOR STANDARD DISTRIBUTIONS

The expectations and variances for the special distributions described in previous sections are as follows:

		Parameters	EXPECTATION	VARIANCE
DISCRETE	$Bernoulli(\theta)$	θ	θ	$\theta(1-\theta)$
	$Binomial(n,\theta)$	n,θ	$n\theta$	$n\theta(1-\theta)$
	$Poisson(\lambda)$	λ	λ	λ
	$Geometric(\theta)$	θ	$\dfrac{1}{\theta}$	$\dfrac{(1-\theta)}{\theta^2}$
	$NegBinomial(n,\theta)$	n,θ	$\dfrac{n}{\theta}$	$\dfrac{n(1-\theta)}{\theta^2}$
CONTINUOUS	$Uniform(a,b)$	a,b	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
	$Exponential(\lambda)$	λ	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
	$Gamma(\alpha,\beta)$	α,β	$\dfrac{\alpha}{\beta}$	$\dfrac{\alpha}{\beta^2}$
	$Beta(\alpha,\beta)$	α,β	$\dfrac{\alpha}{\alpha+\beta}$	$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
	$Normal(\mu,\sigma^2)$	μ,σ^2	μ	σ^2

BERNOULLI

EXPECTATION $\quad \mathrm{E}_{f_X}[\,X\,] = \sum_{x=0}^{1} x\theta^x(1-\theta)^{1-x} = \theta$

VARIANCE $\quad \mathrm{Var}_{f_X}[\,X\,] = \mathrm{E}_{f_X}\left[X^2\right] - \{\mathrm{E}_{f_X}[\,X\,]\}^2 = \sum_{x=0}^{1} x^2\theta^x(1-\theta)^{1-x} - \theta^2 = \theta(1-\theta)$

BINOMIAL

Recall that the $Binomial(n,\theta)$ distribution arises when we form the sum of n independent and identically distributed $Bernoulli\,(\theta)$ random variables, X_i. Hence using the results from the previous section, we have that

EXPECTATION $\quad \mathrm{E}_{f_X}[\,X\,] = \sum_{x=1}^{n} \mathrm{E}_{f_{X_i}}[\,X_i\,] = \sum_{x=1}^{n} \theta = n\theta$

VARIANCE $\quad \mathrm{Var}_{f_X}[\,X\,] = \sum_{x=1}^{n} \mathrm{Var}_{f_{X_i}}[\,X_i] = \sum_{x=1}^{n} \theta\,(1-\theta) = n\theta\,(1-\theta)$

GEOMETRIC

EXPECTATION $\quad \mathrm{E}_{f_X}[\,X\,] = \sum_{x=1}^{n} x(1-\theta)^{x-1}\theta = \theta \sum_{x=1}^{n} \left\{\frac{d}{d\theta}\{-(1-\theta)^x\}\right\} = -\theta\frac{d}{d\theta}\left\{\sum_{x=0}^{\infty}(1-\theta)^x\right\}$

$$= -\theta\frac{d}{d\theta}\left\{\frac{1}{1-(1-\theta)}\right\} = -\theta\,\frac{-1}{\theta^2} = \tfrac{1}{\theta}$$

VARIANCE $\quad \mathrm{Var}_{f_X}[\,X\,] = \dfrac{1-\theta}{\theta^2}$

NEGATIVE BINOMIAL

Recall that the $NegativeBinomial(n,\theta)$ distribution arises when we form the sum of n independent and identically distributed $Geometric\,(\theta)$ random variables, X_i. Hence using the results from the previous section, we have that

$$\text{EXPECTATION} \quad \mathrm{E}_{f_X}[\ X\ ] = \sum_{x=1}^{n} \mathrm{E}_{f_{X_i}}[\ X_i\ ] = \sum_{x=1}^{n} \frac{1}{\theta} = \frac{n}{\theta}$$

$$\text{VARIANCE} \quad \mathrm{Var}_{f_X}[\ X\ ] = \sum_{x=1}^{n} \mathrm{Var}_{f_{X_i}}[\ X_i] = \sum_{x=1}^{n} \frac{1-\theta}{\theta^2} = \frac{n\,(1-\theta)}{\theta^2}$$

POISSON

$$\text{EXPECTATION} \quad \mathrm{E}_{f_X}[\ X\ ] = \sum_{x=0}^{\infty} x\frac{\lambda^x}{x!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda}\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \lambda e^{-\lambda}e^{\lambda} = \lambda$$

For the variance, we first note that

$$\mathrm{Var}_{f_X}[\ X\ ] = \mathrm{E}_{f_X}[\ (X - \{\mathrm{E}_{f_X}[\ X\ ]\})^2\ ] = \mathrm{E}_{f_X}[\ X^2\ ] - \{\mathrm{E}_{f_X}[\ X\ ]\}^2 = \mathrm{E}_{f_X}[\ X(X-1)+X\ ] - \{\mathrm{E}_{f_X}[\ X\ ]\}^2$$

and hence, using the properties of expectation

$$\mathrm{Var}_{f_X}[\ X\ ] = \mathrm{E}_{f_X}[\ X(X-1)\ ] + \mathrm{E}_{f_X}[\ X\ ] - \lambda^2$$

where

$$\mathrm{E}_{f_X}[\ X(X-1)\ ] = \sum_{x=2}^{\infty} x(x-1)\frac{\lambda^x}{x!}e^{-\lambda} = \lambda^2 e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!} = \lambda^2 e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = \lambda^2 e^{-\lambda}e^{\lambda} = \lambda^2$$

so

$$\text{VARIANCE} \quad \mathrm{Var}_{f_X}[\ X\ ] = \lambda^2 + \lambda - \lambda^2 = \lambda$$

EXPONENTIAL

Note first that, using integration by parts, for positive integer value k

$$\begin{aligned} I_k &= \mathrm{E}_{f_X}\left[\ X^k\right] = \int_0^\infty x^k \lambda e^{-\lambda x}\ dx = \left[-x^k e^{-\lambda x}\right]_0^\infty + \int_0^\infty k x^{k-1} e^{-\lambda x}\ dx = 0 + \frac{k}{\lambda}\int_0^\infty x^{k-1}\lambda e^{-\lambda x}\ dx \\ &= \frac{k}{\lambda} I_{k-1} = \frac{k(k-1)}{\lambda^2} I_{k-2} = \dots = \frac{k!}{\lambda^k} I_0 = \frac{k!}{\lambda^k} \qquad \text{as } I_0 = \int_0^\infty \lambda e^{-\lambda x}\ dx = 1. \end{aligned}$$

Hence

$$\text{EXPECTATION} \quad \mathrm{E}_{f_X}[\ X\ ] = \int_0^\infty x\lambda e^{-\lambda x}\ dx = I_1 = \frac{1}{\lambda}$$

$$\text{VARIANCE} \quad \mathrm{Var}_{f_X}[\ X\ ] = \mathrm{E}_{f_X}[\ X^2\ ] - \{\mathrm{E}_{f_X}[\ X\ ]\}^2 == I_2 - \{I_1\}^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

where the integration by parts is straightforward in both cases.

GAMMA

First consider the expectation of X^k for $k = 1, 2, ...$, that is,

$$\mathrm{E}_{f_X}\left[X^k\right] = \int_{-\infty}^{\infty} x^k f_X(x)\,dx = \int_0^{\infty} x^k \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\,dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{k+\alpha-1} e^{-\beta x}\,dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+k)}{\beta^{\alpha+k}}$$

as the integrand in the last integral is proportional to a Gamma pdf with parameters $k+\alpha$ and β. Hence

$$\mathrm{E}_{f_X}[\,X^k\,] = \frac{\alpha(\alpha+1)...(\alpha+k-1)}{\beta^k} \quad \therefore \left\{ \begin{array}{ll} \mathrm{E}_{f_X}[\,X\,] & = \dfrac{\alpha}{\beta} \\ & \\ \mathrm{Var}_{f_X}[\,X\,] & = \mathrm{E}_{f_X}\left[X^2\right] - \left\{\mathrm{E}_{f_X}[X]\right\}^2 = \dfrac{\alpha(\alpha+1)}{\beta^2} - \dfrac{\alpha^2}{\beta^2} = \dfrac{\alpha}{\beta^2} \end{array} \right.$$

BETA

First consider the expectation of X^k for $k = 1, 2, ...$, that is,

$$\begin{aligned} \mathrm{E}_{f_X}\left[X^k\right] &= \int_{-\infty}^{\infty} x^k f_X(x)\,dx = \int_0^1 x^k \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{k+\alpha-1}(1-x)^{\beta-1}\,dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(k+\alpha)\Gamma(\beta)}{\Gamma(k+\alpha+\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(k+\alpha)}{\Gamma(k+\alpha+\beta)} \end{aligned}$$

as the integrand in the last integral is proportional to a Beta pdf with parameters $k+\alpha$ and β. Hence

The expectation/variance calculations for the Beta distribution are reasonably straightforward, and it can be shown that

$$\text{EXPECTATION} \quad \mathrm{E}_{f_X}[X] = \frac{\alpha}{\alpha+\beta}$$

$$\text{VARIANCE} \quad \mathrm{Var}_{f_X}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

NORMAL

The expectation of a random variable is the "centre of mass", that is, a measure of centre of the distribution. If a pdf is symmetric about a point c, then we must have that $\mathrm{E}_{f_X}[X] = c$. The Normal pdf is a symmetric function about μ. Hence the expectation should be μ - we can prove this as follows. If $Z \sim Normal(0,1)$, then

$$\begin{aligned} \mathrm{E}_{f_Z}[Z] &= \int_{-\infty}^{\infty} z \left(\frac{1}{2\pi}\right)^{1/2} \exp\left\{-\frac{z^2}{2}\right\} dz = \left(\frac{1}{2\pi}\right)^{1/2} \left[-\exp\left\{-\frac{z^2}{2}\right\}\right]_{-\infty}^{\infty} = 0 \\ \mathrm{E}_{f_Z}[Z^2] &= \int_{-\infty}^{\infty} z^2 \left(\frac{1}{2\pi}\right)^{1/2} \exp\left\{-\frac{z^2}{2}\right\} dz = \left(\frac{1}{2\pi}\right)^{1/2} \int_{-\infty}^{\infty} z^2 \exp\left\{-\frac{z^2}{2}\right\} dz \\ &= \left(\frac{1}{2\pi}\right)^{1/2} \left[-z\exp\left\{-\frac{z^2}{2}\right\}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \left(\frac{1}{2\pi}\right)^{1/2} \exp\left\{-\frac{z^2}{2}\right\} dz = 0 + 1 = 1 \end{aligned}$$

as the remaining integral is the integral of the standard normal pdf, which must equal 1. Hence

$$\text{EXPECTATION} \quad \mathrm{E}_{f_Z}[Z] = 0$$

$$\text{VARIANCE} \quad \mathrm{Var}_{f_Z}[Z] = \mathrm{E}_{f_Z}\left[Z^2\right] - \left\{\mathrm{E}_{f_Z}[Z]\right\}^2 = 1 - 0 = 0$$

Now, using the result that if $Z \sim Normal(0,1)$ then $X = \mu + \sigma Z \sim Normal\left(\mu, \sigma^2\right)$,

$$\text{EXPECTATION} \quad \mathrm{E}_{f_X}[X] = \mu + \sigma \mathrm{E}_{f_Z}[Z] = \mu$$

$$\text{VARIANCE} \quad \mathrm{Var}_{f_X}[X] = \sigma^2 \mathrm{Var}_{f_Z}[Z] = \sigma^2$$

2.8.3 ENTROPY

For a random variable X with mass or density function f_X, the **entropy** of the distribution is defined by

$$H_{f_X}[X] = \mathrm{E}_{f_X}[-\log f_X(X)] = \begin{cases} -\sum_x \log[f_X(x)]\ f_X(x) & \text{DISCRETE CASE} \\ \\ -\int \log[f_X(x)]\ f_X(x)dx & \text{CONTINUOUS CASE} \end{cases}$$

where log in this case can mean logarithm to any base; typically, $\log_2$ or ln (natural log) are used. One interpretation of the entropy of a distribution is that it measures the "evenness" of the distribution, that is, a distribution with high entropy assigns approximately equal probability to each value of the random variable.

EXAMPLE Consider the discrete uniform distribution

$$f_X(x) = \frac{1}{N} \qquad x = 1, 2, ..., N$$

Then

$$H_{f_X}[X] = -\sum_x \log[f_X(x)]\ f_X(x) = -\sum_{x=1}^{N} \log\left[\frac{1}{N}\right] \times \frac{1}{N} = \log N$$

In fact, the uniform distribution is the distribution for which entropy is maximized.

EXAMPLE Consider the discrete distribution

$$f_X(x) = 1 \qquad x = x_0$$

and zero otherwise (so that $\mathrm{P}[X = x_0] = 1$). Then $H_{f_X}[X] = \log[f_X(x_0)] \times f_X(x_0) = 0$

These two examples illustrate another interpretation for the entropy as an overall measure of *uncertainty*. In the first example, there is the maximum possible uncertainty, whereas in the second example the uncertainty is at a minimum.

RELATIVE ENTROPY

It is also possible to compare two distributions using an entropy measure. Consider two discrete distributions with mass functions f_0 and f_1. Then the **relative entropy** of f_1 with respect to f_0 , $H_{f_0||f_1}$, and the **relative entropy** of f_0 with respect to f_1 , $H_{f_1||f_0}$, are defined by

$$H_{f_0||f_1}[X] = E_{f_0}\left[\log\left\{\frac{f_0(X)}{f_1(X)}\right\}\right] = \sum_x \log\left\{\frac{f_0(x)}{f_1(x)}\right\}\ f_0(x) \qquad H_{f_1||f_0}[X] = E_{f_1}\left[\log\left\{\frac{f_1(X)}{f_0(X)}\right\}\right] = \sum_x \log\left\{\frac{f_1(x)}{f_0(x)}\right\}\ f_1(x)$$

where the sum extends over values of x for which both f_0 and f_1 are non-zero. It is also possible to obtain an overall measure of the difference in entropy terms between the two distributions as the sum of these two measures.

$$H_{f_0,f_1}[X] = H_{f_0||f_1}[X] + H_{f_1||f_0}[X] = \sum_x \log\left\{\frac{f_0(x)}{f_1(x)}\right\}\ f_0(x) + \sum_x \log\left\{\frac{f_1(x)}{f_0(x)}\right\}\ f_1(x)$$

It can be shown that $H_{f_0||f_1}[X], H_{f_1||f_0}[X]$ and hence $H_{f_0,f_1}[X]$ are all non-negative. Furthermore, we can define the **support** for x in favour of f_0 over f_1, denoted $S_{0,1}(x)$ by

$$S_{0,1}(x) = \log\left\{\frac{f_0(x)}{f_1(x)}\right\}$$

with the equivalent definition for $S_{1,0}(x)$ (where $S_{1,0}(x) = -S_{0,1}(x)$) Using this definition, we see that $S_{0,1}(X)$ is a random variable, and using the general expectation definition given at the end of section 2.8.1 we have that the expectation of $S_{0,1}(x)$ is

$$\sum_x S_{0,1}(x)\ f_0(x) = \sum_x \log\left\{\frac{f_0(x)}{f_1(x)}\right\}\ f_0(x) = H_{f_0||f_1}[X]$$

2.9. TRANSFORMATIONS OF RANDOM VARIABLES

Consider a discrete or continuous random variable X with range $\mathbb{X}$ and probability distribution described by mass/pdf f_X, or cdf F_X. Suppose g is a function. Then $Y = g(X)$ is also a random variable as Y and typically we wish to derive the probability distribution of random variable Y; in order to do this, we might consider the inverse transformation g^{-1} from $\mathbb{Y}$ to $\mathbb{X}$ Consider first the cdf of Y, F_Y, evaluated at a point $y \in \mathbb{R}$. We have

$$F_Y(y) = \mathrm{P}[\ Y \le y\ ] = \ \mathrm{P}[\ g(X) \le y\ ] = \begin{cases} \displaystyle\sum_{x \in A_y} f_X(x) & \text{if } X \text{ is discrete} \\ \\ \displaystyle\int_{A_y} f_X(x)\ dx & \text{if } X \text{ is continuous} \end{cases}$$

where $A_y = \{\ x \in \mathbb{X}\ :\ g(x) \le y\}$. Often, the set is A_y is easy to identify for a given y, and this becomes our main objective in the calculation.

EXAMPLE Suppose that $g(x) = \exp x$. Then

$$F_Y(y) = \mathrm{P}[\ Y \le y\ ] = \ \mathrm{P}[\ \exp X \le y\ ] = \mathrm{P}[\ X \le \log y\ ] = F_X(\log y)$$

so that $A_y = \{\ x \in \mathbb{X}\ :\ \exp x \le y\} = \{\ x \in \mathbb{X}\ :\ x \le \log y\}$

EXAMPLE Suppose that $g(x) = ax + b$. Then

$$F_Y(y) = \mathrm{P}[\ Y \le y\ ] = \ \mathrm{P}[\ aX + b \le y\ ] = \mathrm{P}\left[X \le \frac{y-b}{a}\right] = F_X\left(\frac{y-b}{a}\right)$$

so that $A_y = \{\ x \in \mathbb{X}\ :\ ax + b \le y\} = \{\ x \in \mathbb{X}\ :\ x \le (y-b)/a\}$

EXAMPLE Suppose that $g(x) = x^2$. Then

$$F_Y(y) = \mathrm{P}[\ Y \le y\ ] = \ \mathrm{P}\left[X^2 \le y\right] = \mathrm{P}\left[-\sqrt{y} \le X \le \sqrt{y}\right] = F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right)$$

so that $A_y = \left\{\ x \in \mathbb{X}\ :\ x^2 \le y\right\} = \left\{\ x \in \mathbb{X}\ :\ \sqrt{y} \le x \le \sqrt{y}\right\}$

We may be interested in the mass or density function of the newly formed variable Y; in that case we could take the cdf formed above and use it to calculate the mass function/pdf. For example, if $Y = aX + b$ when $a > 0$ then

$$F_Y(y) = F_X\left(\frac{y-b}{a}\right) \qquad \Longrightarrow \qquad f_Y(y) = \frac{d}{dy}\left\{F_X\left(\frac{y-b}{a}\right)\right\} = \frac{d}{dy}\left\{\left(\frac{y-b}{a}\right)\right\} f_Y\left(\frac{y-b}{a}\right) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

using the chain rule for differentiation that says

$$\frac{d}{dx}\left\{g(h(x)\right\} = h'(x)g'(h(x)) \qquad \text{where } g'(x) = \frac{dg(x)}{dx} \qquad h'(x) = \frac{dh(x)}{dx}$$

In the discrete case, it may be easier to consider the mass function directly rather than the cdf. However, for a particular type of transformations, namely 1-1 transformations, it is possible to produce a general transformation result that allows direct calculation of the distribution of the transformed variable.

2.9.3 1-1 TRANSFORMATIONS

The mapping $g(X)$ is a function of X from $\mathbb{X}$ which is 1-1 and onto $\mathbb{Y}$ if,

(i) for each $x \in \mathbb{X}$, there exists one and only one y such that $y = g(x)$, and

(ii) for each $y \in \mathbb{Y}$, there exists an $x \in \mathbb{X}$ such that $g(x) = y$.

Most transformations are 1-1 transformations (the exceptions being transformations involving powers of X, like $g(x) = x^2$, or $g(x) = x(1-x)$. The following result gives the distribution for random variable $Y = g(X)$ when g is 1-1.

THEOREM

Let X be a random variable with mass/density function f_X and support $\mathbb{X}$. Let g be a 1-1 function from $\mathbb{X}$ onto $\mathbb{Y}$ with inverse g^{-1}. Then $Y = g(X)$ is a random variable with support $\mathbb{Y}$ and

Discrete Case : The mass function of random variable Y is given by

$$f_Y(y) = f_X(g^{-1}(y))$$

for $y \in \mathbb{Y} = \{\ y \mid f_Y(y) > 0\ \}$, where x is the unique solution of $y = g(x)$ (so that $x = g^{-1}(y)$).

Continuous Case : The pdf of random variable Y is given by

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dt}\left\{g^{-1}(t)\right\}_{t=y} \right|$$

$y \in \mathbb{Y} = \{\ y \mid f_Y(y) > 0\ \}$where $y = g(x)$, provided that the derivative $\dfrac{d}{dt}\left\{g^{-1}(t)\right\}$ is continuous and non-zero on $\mathbb{Y}$.

PROOF

Discrete case: by direct calculation,

$$f_Y(y) = P[\ Y = y\ ] = P[\ g(X) = y\ ] = P[\ X = g^{-1}(y)\ ] = f_X(g^{-1}(y)) = f_X(x)$$

. where $x = g^{-1}(y)$, and hence $f_Y(y) > 0 \Longleftrightarrow f_X(x) > 0$.

Continuous case: function g is either (I) a monotonic increasing, or (II) a monotonic decreasing function.

Case (I): If g is increasing, then for $x \in \mathbb{X}$ and $y \in \mathbb{Y}$, we have that $g(x) \leq y \Longleftrightarrow x \leq g^{-1}(y)$. Therefore, for $y \in \mathbb{Y}$,

$$F_Y(y) = \mathrm{P}[\ Y \leq y\ ] = \mathrm{P}[\ g(X) \leq y\ ] = \mathrm{P}[\ X \leq g^{-1}(y)\ ] = F_X(g^{-1}(y))$$

and, by differentiation, because g is monotonic increasing,

$$f_Y(y) = f_X(g^{-1}(y)) \frac{d}{dt}\left\{g^{-1}(t)\right\}_{t=y} = f_X(g^{-1}(y)) \left| \frac{d}{dy}\left\{g^{-1}(y)\right\}_{t=y} \right| \qquad \text{as } \frac{d}{dt}\left\{g^{-1}(t)\right\} > 0.$$

Case (II): If g is decreasing, then for $x \in \mathbb{X}$ and $y \in \mathbb{Y}$ we have $g(x) \leq y \Longleftrightarrow x \geq g^{-1}(y)$. Therefore, for $y \in \mathbb{Y}$,

$$F_Y(y) = \mathrm{P}[\ Y \leq y\ ] = \mathrm{P}[\ g(X) \leq y\ ] = \mathrm{P}[\ X \geq g^{-1}(y)\ ] = 1 - F_X(g^{-1}(y))$$

so by differentiation

$$f_Y(y) = -f_X(g^{-1}(y)) \frac{d}{dt}\left\{g^{-1}(t)\right\}_{t=y} = f_X(g^{-1}(y)) \left| \frac{d}{dt}\left\{g^{-1}(t)\right\}_{t=y} \right| \qquad \text{as} \frac{d}{dt}\left\{g^{-1}(t)\right\} < 0.$$

The **Jacobian** of the transformation defined by $g(x) = y$ for $x \in \mathbb{X}$, denoted $J(y)$, is given by

$$J(y) = \left| \frac{d}{dt}\left\{g^{-1}(t)\right\}_{t=y} \right|$$

that is, the absolute value of first derivative of g^{-1} evaluated at $y = g(x)$. Note that the inverse transformation $g^{-1} : \mathbb{Y} \longrightarrow \mathbb{X}$ has Jacobian $1/J(x)$.

LOCATION/SCALE TRANSFORMATIONS

A particular type of 1-1 transformation is a *location/scale* transformation; this transformation take the form

$$Y = \mu + \lambda X$$

where μ and $\lambda > 0$ are two real-parameters. The pdf of the location/scale transformed variable Y was derived above using first principles

$$\begin{aligned} F_Y(y) &= \mathrm{P}[Y \leq y] = \mathrm{P}[\mu + \lambda X \leq y] = \mathrm{P}\left[X \leq \frac{y-\mu}{\lambda}\right] = F_X\left(\frac{y-\mu}{\lambda}\right) \\ \Longrightarrow f_Y(y) &= \frac{d}{dy}\left\{F_X\left(\frac{y-\mu}{\lambda}\right)\right\} = \frac{1}{\lambda} f_X\left(\frac{y-b}{\lambda}\right) \end{aligned}$$

Sometimes we are interested in a *scale* transformation only where $\mu = 0$ in the transformation above.

2.9.4 TRANSFORMATION CONNECTIONS BETWEEN DISTRIBUTIONS

Some of the continuous distributions that we have studied are directly connected by transformations

Distribution of X	Transformation	Distribution of Y
$X \sim Uniform(0,1)$	$Y = -\frac{1}{\lambda}\log X$	$Y \sim Exponential(\lambda)$
$X \sim Gamma(\alpha, 1)$	$Y = X/\beta$	$Y \sim Gamma(\alpha, \beta)$
$X \sim Normal(0,1)$	$Y = \mu + \sigma X$	$Y \sim Normal(\mu, \sigma^2)$
$X \sim Normal(0,1)$	$Y = X^2$	$Y \sim Gamma\left(\frac{1}{2}, \frac{1}{2}\right) \equiv \chi_1$