# MODULE 5

# SOME SPECIAL ABSOLUTELY CONTINUOUS DISTRIBUTIONS

# LECTURE 23

## Topics

## 5.3 BETA DISTRIBUTION

## 5.3 BETA DISTRIBUTION

We will first provide the definition of the beta function.

**Definition 3.1**

The function $B:(0,\infty)\times(0,\infty)\rightarrow(0,\infty)$, defined by,

$$B(x,y)=\int_0^1 t^{x-1}(1-t)^{y-1}dt$$

is called the *beta function*. ■

Clearly the integral

$$\int_0^1 t^{x-1}(1-t)^{y-1}dt$$

converges for $x\geq 1$ and $y\geq 1$. For $x\in(0,1)$ or $y\in(0,1)$ the integral

$$\int_0^1 t^{x-1}(1-t)^{y-1}dt$$

will converge if, and only if, both the integrals

$$\int_0^{\frac{1}{2}} t^{x-1}(1-t)^{y-1}dt \quad \text{and} \int_{\frac{1}{2}}^1 t^{x-1}(1-t)^{y-1}dt$$

converge.

Since, for $0 < x < 1$,

$$\lim_{t\to 0}\frac{t^{x-1}(1-t)^{y-1}}{t^{x-1}} = 1, \qquad \forall y \in \mathbb{R}$$

and the integral

$$\int_0^{\frac{1}{2}} t^{x-1}dt$$

converges, it follows that the integral

$$\int_0^{\frac{1}{2}} t^{x-1}(1-t)^{y-1}dt$$

converges for $(x, y) \in (0,1) \times \mathbb{R}$.

Similarly, for $0 < y < 1$,

$$\lim_{t\to 1}\frac{t^{x-1}(1-t)^{y-1}}{(1-t)^{y-1}} = 1, \quad \forall x \in \mathbb{R}$$

and the integral

$$\int_{\frac{1}{2}}^{1} (1-t)^{y-1}dt$$

converges. Consequently the integral

$$\int_{\frac{1}{2}}^{1} t^{x-1}(1-t)^{y-1}dt$$

converges for $(x, y) \in \mathbb{R} \times (0,1)$.

From the above discussion it follows that the integral

$$\int_0^1 t^{x-1}(1-t)^{y-1}dt$$

converges if $x > 0$ and $y > 0$.

Using the above arguments it can also be seen that the integral

$$\int_0^1 t^{x-1}(1-t)^{y-1}dt$$

diverges if $x \leq 0$ or $y \leq 0$.

Thus the beta function $B: (0,\infty) \times (0,\infty) \to (0,\infty)$ is well defined. For $x > 0$ and $y > 0$, consider

$$\Gamma(x)\,\Gamma(y) = \left(\int_0^\infty e^{-s}\, s^{x-1}ds\right)\left(\int_0^\infty e^{-t}\, t^{y-1}dt\right)$$

$$= \int_0^\infty \int_0^\infty e^{-(s+t)}\, s^{x-1}t^{y-1}ds\, dt.$$

Let us make the transformation $s = uv$ and $t = (1-u)v$ in the above integral. Then the Jacobian of the transformation is

$$J = \begin{vmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{vmatrix}$$

$$= \begin{vmatrix} v & u \\ -v & 1-u \end{vmatrix}$$

$$= v.$$

Also,

$$0 < s < \infty \text{ and } 0 < t < \infty \Longrightarrow 0 < u < 1 \text{ and } v > 0.$$

Therefore,

$$\Gamma(x)\Gamma(y) = \int_0^1 \int_0^\infty e^{-v}(uv)^{x-1}\big((1-u)v\big)^{y-1}|v|dv\, du$$

$$= \int_0^1 \left\{\int_0^\infty e^{-v}v^{x+y-1}dv\right\}u^{x-1}(1-u)^{y-1}du$$

$$= \Gamma(x+y) \int_0^1 u^{x-1}(1-u)^{y-1} du$$

$$= \Gamma(x+y)\, B(x,y)$$

$$\Rightarrow \boxed{B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \int_0^1 t^{x-1}(1-t)^{y-1} dt\,, \ \ x > 0, y > 0.}$$

Note that $B(x,y) = B(y,x), \forall (x,y) \in (0,\infty) \times (0,\infty)$.

**Definition 3.2**

Let $X$ be a random variable of absolutely continuous type and let $a > 0$ and $b > 0$ be real constants. The random variable $X$ is said to follow the *beta distribution* with shape parameter $(a,b)$ (written as $X \sim \mathrm{Be}(a,b)$) if its probability density function is given by

$$f_X(x) = \begin{cases} \dfrac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}. \ \blacksquare$$

Clearly $f_X(x) \geq 0$, $\forall x \in \mathbb{R}$ and

$$\int_0^\infty f_X(x) dx = \frac{1}{B(a,b)} \int_0^1 x^{a-1}(1-x)^{b-1} dx = 1.$$

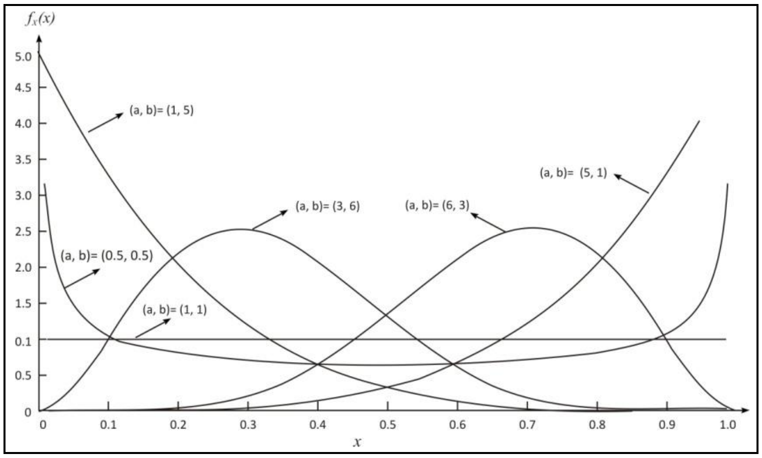


**Figure 3.1.** Plots of p.d.f.s of $\text{Be}(a, b)$ distributions

Note that $\text{Be}(1, 1)$ distribution is nothing but $U(0, 1)$ distribution.

Suppose that $X \sim \text{Be}(a, b)$ distribution, for some positive constants $a$ and $b$. Then, for $r > -a$

$$E(X^r) = \frac{1}{B(a,b)} \int_0^1 x^r x^{a-1} (1-x)^{b-1} dx$$

$$= \frac{1}{B(a,b)} \int_0^1 x^{a+r-1} (1-x)^{b-1} dx$$

$$\text{i.e.,} \boxed{E(X^r) = \frac{B(a+r,b)}{B(a,b)} = \frac{\Gamma(a+r)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+r)}, \qquad r > -a.}$$

Therefore,

$$\boxed{\text{Mean} = \mu_1' = E(X) = \frac{a}{a+b},}$$

$$\mu_2' = E(X^2) = \frac{a(a+1)}{(a+b)(a+b+1)},$$

$$\boxed{\text{Variance} = \mu_2 = E(X^2) - (E(X))^2 = \frac{ab}{(a+b)^2(a+b+1)},}$$

$$\mu_3 = E\left(\left(X - \mu_1'\right)^3\right) = \mu_3' - 3\,\mu_1'\mu_2' + 2\left(\mu_1'\right)^3 = \frac{2(b-a)ab}{(a+b)^3(a+b+1)(a+b+2)},$$

$$\mu_4 = E\left(\left(X-\mu_1'\right)^4\right) = \mu_4' - 4\mu_1'\mu_3' + 6\left(\mu_1'\right)^2\mu_2' - 3\left(\mu_1'\right)^4$$

$$= \frac{3ab(2(b-a)^2+ab)}{(a+b)^4(a+b+1)(a+b+2)(a+b+3)},$$

$$\boxed{\text{Coefficient of skewness} = \beta_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{2(b-a)}{a+b+2}\sqrt{\frac{a+b+1}{ab}},}$$

and

$$\boxed{\text{Kurtosis} = \gamma_1 = \frac{\mu_4}{\mu_2^2} = \frac{3(a+b+1)[2(b-a)^2+ab(a+b+2)]}{ab(a+b+2)(a+b+3)}.}$$

The m.g.f. of $X \sim \mathrm{Be}(a,b)$ is given by

$$M_X(t) = E(e^{tx})$$

$$= \frac{1}{B(a,b)}\int_0^1 e^{tx}x^{a-1}(1-x)^{b-1}dx$$

$$= \frac{1}{B(a,b)}\int_0^1 \left\{\sum_{j=0}^{\infty}\frac{t^j x^j}{j!}\right\} x^{a-1}(1-x)^{b-1}dx$$

$$= \sum_{j=0}^{\infty}\frac{t^j}{j!\,B(a,b)}\int_0^1 x^{a+j-1}(1-x)^{b-1}dx$$

$$= \sum_{j=0}^{\infty}\frac{t^j\,B(a+j,b)}{j!\,B(a,b)}, \qquad t \in \mathbb{R},$$

i, e.,

$$\boxed{M_X(t) = \sum_{j=0}^{\infty}\frac{\Gamma(a+b)\Gamma(a+j)}{\Gamma(a)\Gamma(a+b+j)}\frac{t^j}{j!}, \quad t \in \mathbb{R}.}$$

For $a = b = \alpha\ (>0)$, say and $x \in (0,1)$

$$P(\{X \le x\}) = \frac{1}{B(\alpha,\alpha)}\int_0^x t^{\alpha-1}(1-t)^{\alpha-1}dt$$

$$= \frac{1}{B(\alpha,\alpha)} \int_{1-x}^{1} y^{\alpha-1}(1-y)^{\alpha-1}dy$$

$$= P(\{X \geq 1-x\})$$

$$= P(\{1-X \leq x\}).$$

It follows that

$$P(\{X \leq x\}) = P(\{1-X \leq x\}), \forall x \in \mathbb{R}.$$

Therefore

$$\boxed{X \sim \text{Be}(a,b) \Longrightarrow X \stackrel{d}{=} 1-X \Leftrightarrow X - \frac{1}{2} \stackrel{d}{=} \frac{1}{2} - X.}$$

Thus if $X \sim \text{Be}\,(\alpha,\alpha)$, for some $\alpha > 0$, then the distribution of $X$ is symmetric about $\frac{1}{2}$.

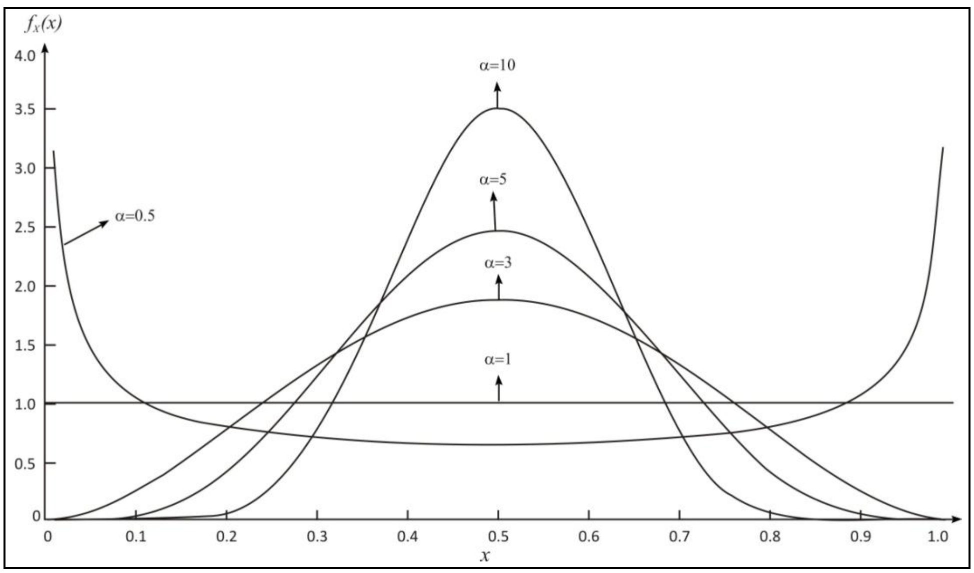


**Figure 3.2.** Plots of p.d.f.s of $\text{Be}(\alpha,\alpha)$ distributions

In the following theorem we establish a relationship between the beta and the binomial probabilities.

**Theorem 3.1**

Let $X \sim \text{Be}(m,n)$, for some positive integers $m$ and $n$. Then, for $x \in (0,1)$,

$$P(\{X \leq x\}) = P(\{Y \geq m\}),$$

where $Y \sim \text{Bin}(m+n-1,x)$. Equivalently

$$\frac{1}{\mathrm{B}(m,n)}\int_0^x t^{m-1}(1-t)^{n-1}dt = \sum_{j=m}^{m+n-1}\binom{m+n-1}{j}x^j(1-x)^{m+n-1-j}, \qquad x\in(0,1).$$

**Proof.** Fix $x \in (0,1)$ and define

$$I_{m,n} = P(\{X \le x\})$$

$$= \frac{1}{\mathrm{B}(m,n)}\int_0^x t^{m-1}(1-t)^{n-1}dt$$

$$= \frac{(m+n-1)!}{(m-1)!\,(n-1)!}\int_0^x t^{m-1}(1-t)^{n-1}dt.$$

On integrating by parts we get

$$I_{m,n} = \frac{(m+n-1)!}{(m-1)!\,(n-1)!}\left[\left\{\frac{t^m(1-t)^{n-1}}{m}\right\}_0^x + \frac{n-1}{m}\int_0^x t^m(1-t)^{n-2}dt\right]$$

$$= \frac{(m+n-1)!}{m!\,(n-1)!}x^m(1-x)^{n-1} + \frac{(m+n-1)!}{m!\,(n-2)!}\int_0^x t^m(1-t)^{n-2}dt$$

$$= \binom{m+n-1}{m}x^m(1-x)^{n-1} + I_{m+1,n-1}$$

$$= \binom{m+n-1}{m}x^m(1-x)^{n-1} + \binom{m+n-1}{m+1}x^{m+1}(1-x)^{n-2} + I_{m+2,n-2}$$

⋮

$$= \sum_{j=m}^{m+n-2}\binom{m+n-1}{j}x^j(1-x)^{m+n-1-j} + I_{m+n-1,1}$$

$$= \sum_{j=m}^{m+n-2}\binom{m+n-1}{j}x^j(1-x)^{m+n-1-j} + \frac{(m+n-1)!}{(m+n-2)!\,0!}\int_0^x t^{m+n-2}dt$$

$$= \sum_{j=m}^{m+n-1}\binom{m+n-1}{j}x^j(1-x)^{m+n-1-j}. \blacksquare$$

**Example 3.1**

Time (in hours) to finish a job follows beta distribution with mean $1/3$ hours and variance $2/63$ hours. Find the probability that the job will be finished in $30$ minutes.

**Solution.** Let $X$ denote the time to finish the job. Then $X \sim \text{Be}(a, b)$, for some $a > 0$ and $b > 0$. We have

$$\text{Mean} = E(X) = \frac{a}{a+b} = \frac{1}{3} \text{ and } \quad \text{Variance} = \text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)} = \frac{2}{63}$$

$$\Rightarrow a = 2 \text{ and } b = 4, \qquad \text{i.e.}, X \sim \text{Be}(2, 4),$$

and therefore the required probability is

$$P\left(\left\{X < \frac{1}{2}\right\}\right) = \frac{1}{B(2,4)} \int_0^{\frac{1}{2}} x(1-x)^3 dx$$

$$= 20 \int_0^{\frac{1}{2}} (x - 3x^2 + 3x^3 - x^4) dx$$

$$= \frac{13}{16}. \blacksquare$$