Section 3.3. The Γ, χ^2, and β Distributions

Note. In this section we define the gamma function using an improper integral. We then use it to define three distributions: the Γ distribution, the χ^2 distribution, and the β distribution. These distributions have a number of applications, some related to lifetimes, failure times, and waiting times.

Definition. The *gamma function* is $\Gamma(a) = \int_0^\infty y^{\alpha-1}e^{-y}\,dy$, where $\alpha > 0$.

Note. We now show that $\Gamma(\alpha)$ actually exists for all $\alpha > 0$; that is, the indefinite integral defining $\Gamma(\alpha)$ is convergent for all $\alpha > 0$. The following computations are largely based on Jia-Ming (Frank) Liou's Calculus 2 webpage (accessed 6/28/2021).

Note 3.3.A. We can show by an inductive application of L'Hoptial's Rule (see my online notes for Calculus 1 on Section 4.5. Indeterminate Forms and L'Hôpital's Rule) that for $n \in \mathbb{N}$ we have $\lim_{y\to\infty} \frac{y^{n-1}}{e^{y/2}} = 0$. So by the definition of "limit as $y \to \infty$," there exists $M > 0$ such that for all $y \geq M$ we have $|y^{n-1}/e^{y/2}| < 1$. So for $y \geq M > 0$ we also have $0 \leq y^{n-1} < e^{y/2}$ when $y > 0$, or $0 \leq y^{n-1}e^{-y} \leq e^{y/2}e^{-y} = e^{-y/2}$. Now

$$\int_0^\infty e^{-y/2}\,dy = \lim_{b\to\infty}\left(\int_0^b e^{-y/2}\,dy\right) = \lim_{b\to\infty}(-2e^{-y/2})|_0^b$$

$$= \lim_{b\to\infty}\left((-2^{-(b)/2}) - (-2e^{-(0)/2})\right) = 2.$$

So by the Direct Comparison Test (see my online notes for Calculus 2 [MATH 1920] on Section 8.7. Improper Integrals; see Theorem 2), $\int_0^\infty y^{n-1}e^{-y}\,dy$ converges when $n \in \mathbb{N}$. So we need to consider the integral when $n \in \mathbb{N}$ is replace with $\alpha > 0$. With $\lfloor \alpha \rfloor$ as the greatest integer function (or the "integer floor function") then for $\alpha \geq 1$ we have $\lfloor \alpha \rfloor \leq \alpha < \lfloor \alpha \rfloor + 1$ or $\alpha - 1 \leq \lfloor \alpha \rfloor \in \mathbb{N}$. So for $x > 0$ we have $0 \leq y^{\alpha-1}e^{-y} \leq x^{\lfloor \alpha \rfloor}e^{-y}$. So $\lfloor \alpha \rfloor \in \mathbb{N}$ and so $\int_0^\infty y^{\alpha-1}e^{-y}\,dy$ converges (by the Direct Comparison Test) when $\alpha \geq 1$.

Note 3.3.B. We now consider $0 < \alpha < 1$. Then $-1 < \alpha - 1 < 0$ and so for $y \geq 1$ we have $0 < y^{\alpha-1} \leq 1$. This implies that $0 < y^{\alpha-1}e^{-y} \leq e^{-y}$ for $y \geq 1$. So by the Direct Comparison Test,

$$\int_1^\infty y^{\alpha-1}e^{-y}\,dy \leq \int_1^\infty e^{-y}\,dy = \lim_{b\to\infty}\left(\int_1^b e^{-y}\,dy\right)$$

$$= \lim_{b\to\infty}(-e^{-y})\big|_1^b = \lim_{b\to\infty}\left((-e^{-b}) - (-e^{-1})\right) = 1/e$$

and hence

$$\int_1^\infty y^{\alpha-1}e^{-y}\,dy \text{ is convergent.} \qquad (*)$$

Finally, we show that $\int_0^1 y^{\alpha-1}e^{-y}\,dy$ converges (notice that

$$\lim_{y\to 0}(y^{\alpha-1}e^{-y}) = \lim_{y\to\infty}\frac{e^{-y}}{y^{1-\alpha}} = \infty$$

since $0 < \alpha < 1$ and so $0 < 1 - \alpha < 1$, and so this is in fact an improper integral). Now $0 < y^{\alpha-1}e^{-y} \leq y^{\alpha-1}e$ for $y \geq 0$, so by the Direct Comparison Test

$$\int_0^1 y^{\alpha-1}e^{-y}\,dy \leq \int_0^1 y^{\alpha-1}e\,dy = \lim_{a\to 0^+}\left(e\frac{y^\alpha}{\alpha}\right)\bigg|_a^1$$

$$= \lim_{a\to 0^+}\left(e\frac{(1)^n}{\alpha} - e\frac{(a)^\alpha}{\alpha}\right) = \frac{e}{\alpha} - e\frac{(0)^\alpha}{\alpha} = \frac{e}{\alpha}$$

and so

$$\int_0^1 y^{\alpha-1}e^{-y}\,dy \text{ converges when } 0 < \alpha < 1. \qquad (**)$$

Therefore, by $(*)$ and $(**)$, we have that

$$\int_0^\infty y^{\alpha-1}e^{-y}\,dy = \int_0^1 y^{\alpha-1}e^{-y}\,dy + \int_1^\infty y^{\alpha-1}e^{-y}\,dy$$

is convergent ("exists"). Hence $\Gamma(\alpha)$ is defined for all $0 < \alpha < 1$. Combining this with Note 3.3.A, we see that $\Gamma(\alpha)$ is defined for all $\alpha > 0$.

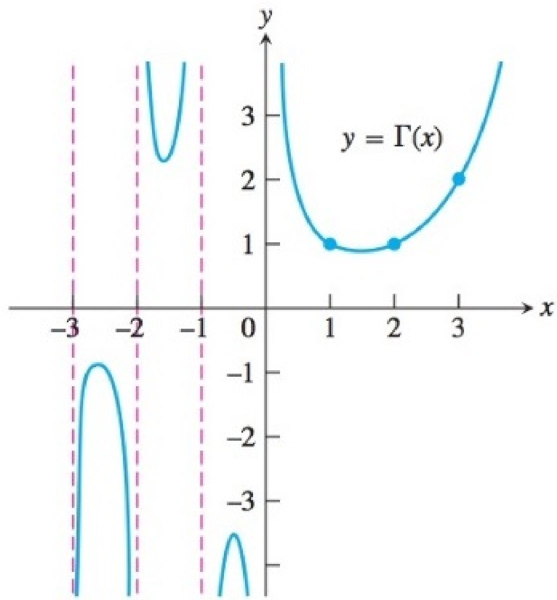
Note. We have

$$\Gamma(1) = \int_0^\infty e^{(1)-1}e^{-y}\,dy = \int_0^\infty e^{-y}\,dy = 1.$$

We can use Integration by Parts to show for $\alpha > 1$ that $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ (see *Thomas' Calculus, Early Transcendentals*, 14th Edition, "Chapter 8. Techniques of Integration," "Additional and Advanced Exercises" Number 43). So by induction we have $\Gamma(n) = (n-1)!$ where $n \in \mathbb{N}$. So the Γ function generalizes the factorial function by extending it from $\mathbb{N}$ to $(0,\infty)$. In fact, the Γ function can be extended to the complex plane where it is defined, except that it has simple poles at $0, -1, -2, \ldots$. Formally, it is defined as

$$\Gamma(z) = \frac{e^{-\gamma z}}{z}\prod_{n=1}^\infty \left(1+\frac{z}{n}\right)^{-1} e^{z/n}$$

where γ is a constant (called the *Euler constant*) such that $\Gamma(1) = 1$. See my online notes for Complex Analysis 1 and 2 (MATH 5510/5520) on Section VII.7. The Gamma Function (see in particular Definition VII.7.2 and Theorem VII.7.15). A graph of the Γ function can be found in *Thomas' Calculus, Early Transcendentals* (in the exercise referenced above):



Definition. Continuous random variable X has a Γ-*distribution* with parameters $\alpha > 0$ and $\beta > 0$, if its probability density function is

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & \text{for } 0 < x < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

We denote this by saying that X has a $\Gamma(\alpha, \beta)$ distribution.

Note. We need to confirm that f actually integrates to 1 over $[0, \infty)$ to insure that it is a probability density function. We have

$$\int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}\, dx \qquad [\text{ let } z = x/\beta \text{ and } dx = 1/\beta\, dx]$$
$$= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} (\beta z)^{\alpha-1} e^{-z} \beta\, dz$$
$$= \frac{\beta^\alpha}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty z^{\alpha-1} e^{-z}\, dz = \frac{1}{\Gamma(\alpha)} \Gamma(\alpha) = 1.$$

Note 3.3.C. The moment generating function of the Γ distribution is

$$\begin{aligned}
M(t) &= E[e^{tX}] = \int_0^\infty e^{tx}\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}\,dx \\
&= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x(1-\beta t)/\beta}\,dx \\
&\quad \text{let } y = x(1-\beta t)/\beta \text{ where } t < 1/\beta, \text{ or } x = \beta y/(1-\beta t), \\
&\quad \text{and } dx = \beta/(1-\beta t)\,dy \\
&= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha}\left(\frac{\beta y}{1-\beta t}\right)^{\alpha-1} e^{-y}\frac{\beta}{1-\beta t}\,dx \\
&= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha}\left(\frac{\beta y}{1-\beta t}\right)^{\alpha-1} e^{-y}\frac{\beta}{1-\beta t}\,dx \\
&= \frac{1}{\Gamma(\alpha)(1-\beta t)^\alpha}\int_0^\infty y^{\alpha-1}e^{-y}\,dy \\
&= \frac{1}{\Gamma(\alpha)(1-\beta t)^\alpha}\Gamma(\alpha) = \frac{1}{(1-\beta t)^\alpha} \text{ for } t < 1/\beta.
\end{aligned}$$

Then

$$M'(t) = \frac{(-\alpha)(-\beta)}{(1-\beta t)^{\alpha+1}} = \frac{\alpha\beta}{(1-\beta t)^{\alpha+1}}$$

and

$$M''(t) = \frac{(\alpha\beta)(-(\alpha+1))(-\beta)}{(1-\beta t)^{\alpha+2}} = \frac{\alpha(\alpha+1)\beta^2}{(1-\beta t)^{\alpha+2}}.$$

By Note 1.9.B, the mean of the Γ distribution is $\mu = M'(0) = \alpha\beta$ and the variance is

$$\sigma^2 = M''(0) - \mu^2 = \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2 = \alpha\beta^2.$$

Note. Figure 3.3.1 gives six probability density functions for different values of α and β (below).

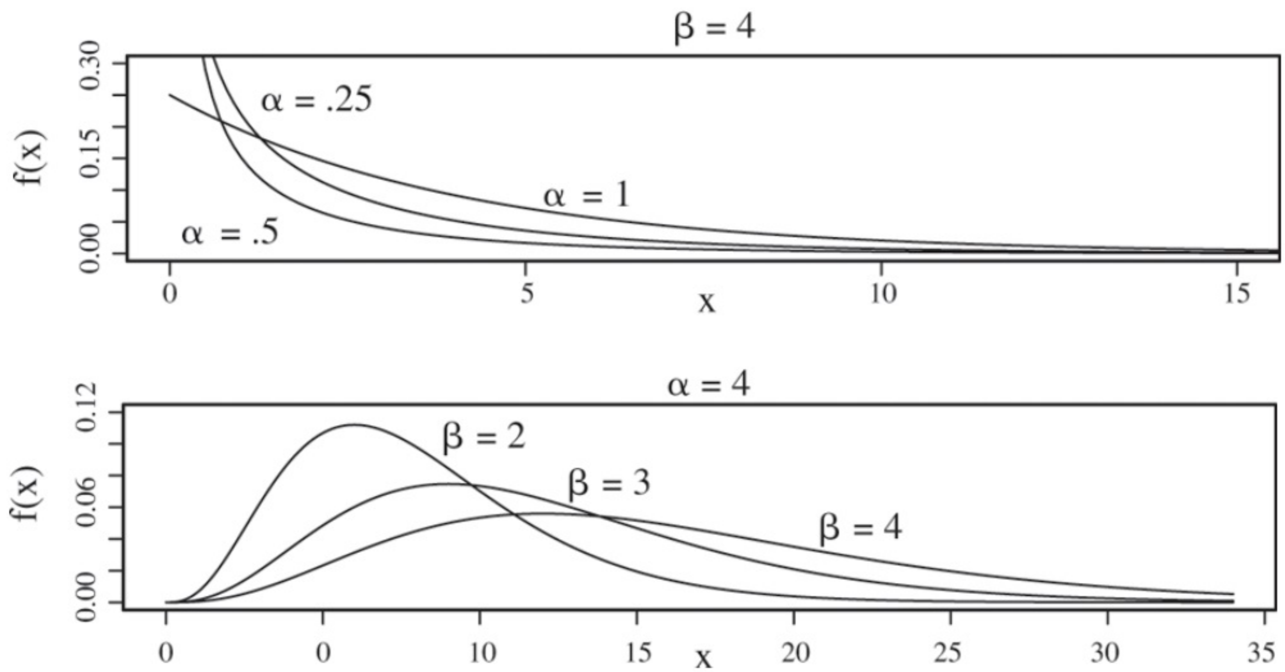


Figure 3.3.1

The appeal of Γ-distribution in applications results from the many different shapes they can take by varying the values of α and β.

Note. Let X denote the time until the failure of a device, and suppose the probability density function is $f(x)$ and the cumulative distribution function is (differentiable) $F(x)$. The "hazard function" of X can be helpful in finding $f(x)$. let x be in the support of X. Suppose the device has not failed at time x; that is, suppose $X > x$. The "rate of failure" as a function of time x over the interval x to $x + \Delta$ satisfies

$$r(x) \approx \frac{P(x \leq x < x + \Delta \mid X \geq x)}{\Delta}$$

and

$$r(x) = \lim_{\Delta \to 0} \frac{P(x \leq x < x + \Delta \mid X \geq x)}{\Delta} = \frac{P(x \leq x < x + \Delta)}{P(X \geq x)} \frac{1}{\Delta}$$

$$= \frac{1}{1-F(x)} \lim_{\Delta \to 0} \frac{P(x \le X < x+\Delta)}{\Delta} = \frac{F'(x)}{1-F(x)} = f(x)1-F(x),$$

since $F'(x) = f(x)$ by Note 1.7.A. This function $r(x)$ is the *hazard function* of X at x. So

$$r(x) = -\frac{d}{dx}[\log(1-F(x)] \text{ or } \log(1-F(x)) \in -\int r(x)\,dx$$

(where we treat the indefinite integral as the set of antiderivatives of the integrand), of

$$1 - F(x) \in e^{-\int r(x)\,dx} \text{ so that } 1 - F(x) = e^{-R(x)}e^C$$

for some $R(x)$ where $R'(x) = r(x)$ and C is some constant. With the support of X as $(0,\infty$ then we take $F(0) = 0$ as the boundary condition that determines constant C.

Note/Definition. If the hazard rate is constant, say $r(x) = 1/\beta$ for some $\beta > 0$, then an antiderivative of r is $R(x) = x/\beta$ and so $1 - F(x) = e^{-x/\beta}e^C$. Since $F(0) = 0$ then $e^C = 1$ so that $F(x) = 1 - e^{-x/\beta}$ and

$$f(x) = F'(x) = \begin{cases} \frac{1}{\beta}e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

This is the $\Gamma(\alpha,\beta) = \Gamma(1,\beta)$ distribution. It is also called the *exponential distribution* with parameter $1/\beta$.

Note. The next result shows that a sum of $\Gamma(\alpha,\beta)$ distributions is additive in the first variable (i.e., the α).

Theorem 3.3.1. Let $X_1, X_2, \ldots, X_n$ be independent random variables. Suppose, for $i = 1, 2, \ldots, n$ that X_i has a $\Gamma(\alpha_i, \beta)$ distribution. Let $Y = \sum_{i=1}^n X_i$. Then Y has a $\Gamma(\sum_{i=1}^n \alpha_i, \beta)$ distribution.

Note 3.3.D. The Γ distribution also arises in Poisson processes. For $t > 0$, the X_t denote the number of events that occur in the interval $(0, t]$, and assume that X_t satisfies the axioms of a Poisson processes given in the previous section. Let k be a fixed positive integer and define the (continuous) random variable W_k to be the "waiting time" until the kth event occurs. The range of W_k is $(0, \infty)$. For any $w > 0$, we have $W_k > w$ (that is, it takes longer than time w for k occurrences of the event of interest) if and only if $X_w \le k - 1$ (that is, at most $k - 1$ events have occurred at time w). Since X_t has a Poisson distribution then

$$P(W_k > w) = P(X_w \le k-1) = \sum_{x=0}^{k-1} P(X_w - x) = \sum_{x=0}^{k-1} \frac{(\lambda w)^x e^{-\lambda w}}{x!}.$$

Now $\displaystyle\int_{\lambda w}^{\infty} \frac{z^{k-1}e^{-z}}{(k-1)!}\,dz = \sum_{x=0}^{k-1} \frac{(\lambda w)^x e^{-\lambda w}}{x!}$ by Exercise 3.3.5. So for $w > 0$, the cumulative distribution function of W_k is

$$\begin{aligned}
f_{W_k}(w) &= 1 - P(W_k > w) = 1 - \sum_{x=0}^{k-1} \frac{(\lambda w)^x e^{-\lambda w}}{x!} \\
&= 1 - \int_{\lambda w}^{\infty} \frac{z^{k-1}e^{-z}}{(k-1)!}\,dz \\
&= 1 - \int \lambda w^{\infty} \frac{z^{k-1}e^{-z}}{\Gamma(k)}\,dz \text{ since } \Gamma(k) = (k-1)! \text{ for } k \in \mathbb{N} \\
&= \frac{\Gamma(k)}{\Gamma(k)} = \frac{1}{\Gamma(k)} \infty)\lambda w^{\infty} z^{k-1}e^{-z}\,dz \\
&= \frac{1}{\Gamma(k)}\left(\int_0^{\infty} z^{k-1}e^{-z}\,dz - \int_{\lambda w}^{\infty} z^{k-1}e^{-z}\,dz\right)
\end{aligned}$$

$$\text{since } \Gamma(k) = \int_0^\infty z^{k-1}e^{-z}\,dz \text{ by definition}$$
$$= \frac{1}{\Gamma(k)}\int_0^{\lambda w} z^{k-1}e^{-z}\,dz,$$

and $F_{W_k}(w) = 0$ for $w \le 0$. With the substitution $z = \lambda y$ and $dz = \lambda\,dy$ in the last integral, we have

$$F_{W_k}(w) = \frac{1}{\Gamma(k)}\int_0^{\lambda w} z^{k-1}e^{-z}\,dz = \int_0^w \frac{\lambda^k y^{k-1}e^{-\lambda y}}{\Gamma(k)}\,dy$$

for $w > 0$. So the probability density function of W_k is

$$\begin{aligned} F_{W_k}(w) &= F'_{W_k}(w) \text{ by Note 1.7.A} \\ &= \frac{d}{dw}\left[\int_0^w \frac{\lambda^k y^{k-1}e^{-\lambda y}}{\Gamma(k)}\,dy\right] = \frac{\lambda^k w^{k-1}e^{-\lambda w}}{\Gamma(k)} \end{aligned}$$

for $w > 0$ and $f_{W_k}(w) = 0$ for $w \le 0$. Since the Γ-distribution, $\Gamma(\alpha,\beta)$, has probability density function

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta} & \text{for } 0 < x < \infty \\ 0 & \text{elsewhere,} \end{cases}$$

then we see that the waiting item until the kth occurrence, W_k, has the Γ-distribution $\Gamma(\alpha,\beta) = \Gamma(k, 1/\lambda)$. In this way, we see the role that a Γ-distribution can play in a Poisson process.

Note. Continuing the previous example on the waiting itme of a Poisson process, if we let T_1 be the waiting time until the first event occurs (so $T_1 = W_1$) then the probability density function of T_1 is

$$f_{T_1}(w) = \Gamma(1, 1/\lambda) = \begin{cases} \lambda e^{-\lambda w} & \text{for } 0 < w < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

So the mean of T_1 is $\alpha\beta = 1/\lambda$ and the mean of X_1 is λ. That is, we expect λ events to occur in a unit of time and we *expect* the first event to occur at time $1/\lambda$. With T_i as the time between the occurrence of event $(i-1)$ and event i, we have that T_i also has a $\Gamma(1, 1/\lambda)$ distribution (this follows from Axioms 1(a) and 2(a) of a Poisson process). Since $T_1, T_2, \ldots$ are independent (by Axiom 3) then the waiting time until the kth event satisfies $W_k = T_1 + T_2 + \cdots T_k$. So by Theorem 3.3.1, W_k has a $\Gamma(k, 1/\lambda)$ distribution (as argued above).

Definition. The distribution $\Gamma(\alpha, \beta) = \Gamma(r/2, 2)$ where $r > 0$ is the *chi-square distribution*, denoted χ^2-distribution, and any probability density function which is the same as the probability density function of $\Gamma(r/2, 2)$ (stated below) is a *chi-square probability function*. Parameter r is the *degrees of freedom* of the χ^2-distribution. The χ^2-distribution with r degrees of freedom is denoted $\chi^2(r)$.

Note. If random variable X has a χ^2-distribution (i.e., a $\Gamma(r/2, 2)$ distribution), then the probability density function is

$$f(x) = \begin{cases} \frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} e^{-x/2} & \text{for } 0 < x < \infty \\ 0 & \text{elsewhere} \end{cases}$$

and the moment generating function is $M(t) = \dfrac{1}{(1-2t)^{r/2}}$ for $t < 1/2$. By Note 3.3.C implies that the mean is $M'(0) = \alpha\beta = r$ and the variance is $\sigma^2 = \alpha\beta^2 = 2r$.

Theorem 3.3.2. Let X have a $\chi^2(r)$ distribution. If $k > -r/2$ then $E(X^k)$ exists and is

$$E(X^k) = \frac{2^k\Gamma(r/2+k)}{\Gamma(r/2)} \text{ if } k > -r/2.$$

Note. The kth moment of the distribution of X is (by definition; see Section 1.9. Some Special Expectations) $E(X^k)$. So Theorem 3.3.2 (since for $k \in \mathbb{N}$ we have $k > -r/2$) the kth moment of the χ^2-distribution is $E(X^k) = 2^k\Gamma(r/2+k)/\Gamma(r/2)$.

Example 3.3.4. Let X have a Γ-distribution with $\alpha = r/2$ (where $4 \in \mathbb{N}$) and $\beta > 0$. Define random variable $Y = 2X/\beta$. The moment generating function of Y is

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = E(e^{2tX/\beta}) \\ &= (1-\beta(2t/\beta))^{-r/2} \text{ by Note 3.3.C} \\ &= (1-2y)^{-r/2}. \end{aligned}$$

Notice that this is the moment generating function of a χ^2-distribution (a $\Gamma(r/2, 2)$ distribution) as shown above. By the uniqueness of the moment generating function (Theorem 1.9.2), we see that Y has a $\chi^2(r)$-distribution.

Corollary 3.3.1. Let $X_1, X_2, \ldots, X_n$ be independent random variables. Suppose X_i has a $\chi^2(r_i)$ distribution for $i = 1, 2, \ldots, n$. Let $Y = \sum_{i=1}^n X_i$. Then Y has a $\chi^2(\sum_{i=1}^n r_i)$-distribution.

Note. The support of any Γ-distribution is the unbounded interval $(0, \infty)$. We now seek a distribution X with support a bounded interval (a, b). Without loss of generality, we may consider random variable Y with support $(0, 1)$ since then $Y = (X-a)/(b-a)$ (or $X = (b-a)Y+a$). An example of such a distribution is a β-distribution. We approach the β-distributions by considering a pair of independent Γ random variables.

Note 3.3.E. Let X_1 and X_2 be two independent random variables that have Γ distributions and have the joint probability density function

$$h(x_1, x_2) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} x_1^{\alpha-1} x_2^{\beta-1} e^{-x_1-x_2} \text{ for } 0 < x_1 < \infty \text{ and } 0 < x_2 < \infty,$$

and $h(x_1, x_2) = 0$ elsewhere; we also require $\alpha > 0$ and $\beta > 0$. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1/(X_1 + X_2)$. We now show that Y_1 and Y_2 are independent (we will choose the marginal distribution of Y_2 as the β distribution). The support of $\mathcal{S}$ of h is the open first quadrant of the x_1x_2-plane. Introduce u_1 and u_2 mapping $\mathcal{S}$ into $\mathbb{R}$ as

$$y_1 = u_1(x_1, x_2) = x_1 + x_2 \text{ and } y_2 = u_2(x_1, x_2) = x_1/(x_1 + x_2).$$

Notice that

$$x_1 = (x_1 + x_2)y_2 = y_1y_2 \text{ and } x_2 = y_1 - x_1 = y_1 - y_1y_2 = y_1(1 - y_2).$$

With

$$x_1 = v_1(y_1, y_2) = y_1y_2 \text{ and } x_2 = v_2(y_1, y_2) = y_1(1 - y_2)$$

we have the Jacobian (or "Jacobian determinant"; see my online Calculus 3 notes on Section 15.8. Substitution in Multiple Integrals)

$$J(x_1,x_2) = J(v_1,v_2) = \begin{vmatrix} \frac{\partial v_1}{\partial y_1} & \frac{\partial v_1}{\partial y_2} \\ \frac{\partial v_2}{\partial y_1} & \frac{\partial v_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

$$= \begin{vmatrix} y_2 & y_1 \\ 1-y_2 & -y_1 \end{vmatrix} = (y_2)(-y_1) - (y_1)(1-y_2) = -y_1 \neq 0 \text{ on } \mathcal{S}.$$

So the transformation $(x_1,x_2) \mapsto (y_1,y_2)$ is one-to-one (injective; this is due to the nonzero Jacobian) mapping of $\mathcal{S}$ onto $\mathcal{T} = \{(y_1,y_2) \mid 0 < y_1 < \infty, 0 < y_2 < 1\}$ is the y_1y_2-plane. Now for $R \subset \mathcal{S}$ a region in the x_1x_2-plane and $G \subset \mathcal{T}$ a region in the y_1y_2-plane such that the transformation $(x_1,x_2) \mapsto (y_1,y_2)$ maps R onto G then

$$\iint_R h(x_1,x_2)\,dx_1\,dx_2 = \iint_G h(v_1(y_1,y_2), v_2(y_1,y_2))|J(v_1,v_2)|\,dv_1\,dv_2$$

provided h, v_1, and v_2 have continuous partial derivatives and $J(v_1,v_2)$ is zero only at isolated points (again, see my online Calculus 3 notes on Section 15.8. Substitution in Multiple Integrals). So we get the probability density function of Y_1 and Y_2 can be obtained form the joint probability density function of X_1 and X_2 by replacing x_1 and y_1y_2, x_2 with $y_1(1-y_2)$, x_1+x_2 with y_1, and introducing $|J(v_1,v_2)| = |-y_1| = y_1$. This gives the joint probability density function of Y_1 and Y_2 on its support of

$$g(y_1,y_2) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)}(y_1y_2)^{\alpha-1}(y_1(1-y_2))^{\beta-1}e^{-y_1}|-y_1|$$

$$= \begin{cases} \frac{y_2^{\alpha-1}(1-y_2)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} y_1^{\alpha+\beta-1}e^{-y_1} & \text{for } 0 < y_1 < \infty, 0 < y_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Since g can be written as a product of a nonnegative function of y_1 with a nonnegative function of y_2, then by Theorem 2.4.1 we have that Y_1 and Y_2 are independent. The marginal probability density function of Y_2 (by Note 2.1.C) is

$$g_2(y_2) = \int_0^\infty \frac{y_2^{\alpha-1}(1-y_2)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} y_1^{\alpha-\beta-1} e^{-y_1}\, dy_1 = \frac{y_2^{\alpha-1}(1-y_2)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty y_1^{\alpha+\beta-1} e^{-y_1}\, dy_1$$

$$= \frac{y_2^{\alpha-1}(1-y_2)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}\Gamma(\alpha+\beta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y_2^{\alpha-1}(1-y_2)^{\beta-1} & \text{for } 0 < y_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

As observed above, Y_1 and Y_2 are independent, so by Definition 2.4.1 the joint probability density function and the marginal probability density functions are related as $g(y_1, y_2) = g_1(y_1)g_2(y_2)$ so it must be that the probability density function of Y_1 is

$$g_1(y_1) = \begin{cases} \frac{1}{\Gamma(\alpha+\beta)} y_1^{\alpha+\beta-1} e^{-y_1} & \text{for} 0 < y_1 < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Notice that $g_1(y_1)$ is the $\Gamma(\alpha+\beta, 1)$ distribution. But it is $g_1(y_2)$ that we are interested in.

Definition. A random variable Y with probability density function

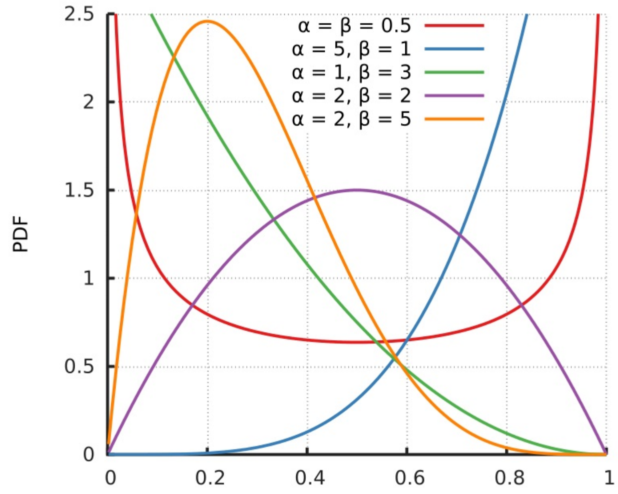
$$g(y) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} & \text{for } 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

is a *beta distribution* (or "β distribution") with parameters α and β.

Note. In Exercise 3.3.A, it is to be shown that the mean and variance of a β distribution with parameters α and β are

$$\mu = \frac{\alpha}{\alpha+\beta} \text{ and } \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}.$$

Note. Here is the shape of the beta distribution for various values of α and β. This image is form the Wikipedia webpage on the beta distribution.



Example 3.3.6. (Dirichlet Distribution) Let $X_1, X_2, \ldots, X_{k+1}$ be independent random variables, each having a Γ distribution with $\beta = 1$. The joint probability density function is then

$$h(x_1, x_2, \ldots, x_{k+1}) = \begin{cases} \prod_{i=1} k+1 \frac{1}{\Gamma(\alpha_i)} x_i^{\alpha_i - 1} e^{-x_i} & \\ 0 & \text{elsewhere.} \end{cases}$$

Let $Y_i = \dfrac{X_i}{X_1 + X_2 + \cdots + X_{k=1}}$ for $i = 1, 2, \ldots, k$, and $Y_{k+1} = X_1 + X_2 + \cdots + X_{k+1}$ denote $k+1$ new random variables. Now the transformation $(x_1, x_2, \ldots, x_{k+1}) \mapsto (y_1, y_2, \ldots, y_{k+1})$ maps set

$$\mathcal{A} = \{(x_1, x_2, \ldots, x_{k+1}) \mid 0 < x_i < \infty \text{ for } i = 1, 2, \ldots, k+1\}$$

onto set

$$\mathcal{B} = \{(y_1, y_2, \ldots, y_{k+1}) \mid 0 < y_i < 1 \text{ for } i = 1, 2, \ldots, k,$$

$$y_1 + y_2 + \cdots y_k < 1, \text{ and } 0 < y_{k+1} < \infty\}.$$

Since $y_{k+1} = x_1 + x_2 + \cdots + x_{k+1}$ then $x_i = y_i y_{k+1}$ for $i = 1, 2, \ldots, k$ and

$$x_{k+1} = y_{k+1} - x_1 - x_2 - \cdots - x_k = y_{k+1} - y_1 y_{k+1} - y_2 y_{k+1} - \cdots - y_k y_{k+1}$$

$$= y_{k+1}(1 - y_1 - y_2 - \cdots - y_k).$$

With $x_i = v_i(y_1, y_2, \ldots, y_{k+1}) = y_i y_{k+1}$ for $i = 1, 2, \ldots, k$ and

$$x_{k+1} = v_{k+1}(y_1, y_2, \ldots, y_{k+1}) = y_{k+1}(1 - y_1 - y_2 - \cdots - y_k),$$

we have the Jacobian

$$\begin{aligned} J(x_1, x_2, \ldots, x_{k+1}) &= J(v_1, v_2, \ldots, v_{k+1}) = \begin{vmatrix} \frac{\partial v_1}{\partial y_1} & \frac{\partial v_1}{\partial y_2} & \cdots & \frac{\partial v_1}{\partial y_{k+1}} \\ \frac{\partial v_2}{\partial y_1} & \frac{\partial v_2}{\partial y_2} & \cdots & \frac{\partial v_2}{\partial y_{k+1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial v_{k+1}}{\partial y_1} & \frac{\partial v_{k+1}}{\partial y_2} & \cdots & \frac{\partial v_{k+1}}{\partial y_{k+1}} \end{vmatrix} \\ &= \begin{vmatrix} y_{k+1} & 0 & \cdots & 0 & y_1 \\ 0 & y_{k+1} & \cdots & 0 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & y_{k+1} & y_k \\ -y_{k+1} & -y_{k+1} & \cdots & -y_{k+1} & (1 - y_1 - y_2 - \cdots - y_k) \end{vmatrix} \\ &= \begin{vmatrix} y_{k+1} & 0 & \cdots & 0 & y_1 \\ 0 & y_{k+1} & \cdots & 0 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & y_{k+1} & y_k \\ 0 & 0 & \cdots & 0 & 1 \end{vmatrix} \quad \text{(by adding each of rows 1 through } k \text{ to row } k+1\text{)} \\ &= y_{k+1}^k \neq 0. \end{aligned}$$

As in Note 3.3.E, the joint probability density function of $Y_1, Y_2, \ldots, T_{k+1}$ can be obtained from the joint probability density function of $X_1, X_2, \ldots, X_{k+1}$ by replacing x_i by $y_i y_{k+1}$ for $i = 1, 2, \ldots, k$, by replacing x_{k+1} by $y_{k+1}(1 - y_1 - y_2 - \cdots y_k)$, and by introducing a factor of $|J(v_1, v_2, \ldots, v_{k+1})| = |y_{k+1}^k| = y_{k+1}^k$. This gives the joint probability density function of $Y_1, Y_2, \ldots, Y_{k+1}$ on its support $\mathcal{B}$ is

$$\begin{aligned}&\{(y_1 y_{k+1})^{\alpha_1 - 1}(y_2 y_{k+1})^{\alpha_2 - 1} \cdots (y_k y_{k+1})^{\alpha_k - 1} y_{k+1}(1 - y_1 - y_2 - \cdots - y_k)^{\alpha_{k+1} - 1}\\ &\times e^{-y_1 y_{k+1}} e^{-y_2 y_{k+1}} \cdots e^{-y_k t_{k+1}} e^{-y_{k+1}(1 - y_1 - y_2 - \cdots - y_k)} y_{k+1}^k\}/(\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_{k+1}))\\ &= \frac{y_{k+1}^{\alpha_1 + \alpha_2 + \cdots + \alpha_{k+1} - 1} y_1^{\alpha_1 - 1} y_2^{\alpha_2 - 1} \cdots y_k^{\alpha_k - 1}(1 - y_1 - y_2 - \cdots - y_k)^{\alpha_{k+1} - 1} e^{-y_{k+1}}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_{k+1})};\end{aligned}$$

and the joint probability density function is 0 off of $\mathcal{B}$. Integrating out y_{k+1} and using the fact that

$$\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_{k+1}) = \int_0^\infty y_{k+1}^{\alpha_1 + \alpha_2 + \cdots + \alpha_{k+1} - 1} e^{y_{k+1}}\, dy_{k+1}$$

we have the joint probability density function of $Y_1, Y_2, \ldots, Y_k$ as

$$g(y_1, y_2, \ldots, y_k) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_{k+1})}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_{k+1})} y_1^{\alpha_1 - 1} y_2^{\alpha_2 - 1} \cdots y_k^{\alpha_k - 1}(1 - y_1 - y_2 - \cdots - y_K)^{\alpha_{k+1} - 1}$$

where $0 < y_i < 1$ for $i = 1, 2, \ldots, k$ and $y_1 + y_2 + \cdots + y_k < 1$. Also $g(y_1, y_2, \ldots, y_k) = 0$ elsewhere. This is the Dirichlet probability density function.

Definition. Random variables $Y_1, Y_2, \ldots, Y_k$ that have a joint probability density function of

$$g(y_1, y_2, \ldots, y_k) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_{k+1})}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_{k+1})} y_1^{\alpha_1 - 1} y_2^{\alpha_2 - 1} \cdots y_k^{\alpha_k - 1}(1 - y_1 - y_2 - \cdots - y_K)^{\alpha_{k+1} - 1}$$

where $0 < y_i < 1$ for $i = 1, 2, \ldots, k$ and $y_1 + y_2 + \cdots + y_k < 1$, and we have $g(y_1, y_2, \ldots, y_k) = 0$ elsewhere, have the *Dirichlet probability density function* with parameters $\alpha_1, \alpha_2, \ldots, \alpha_{k+1}$ where $\alpha_i > 0$ for $i = 1, 2, \ldots, k + 1$.

Note. If $k = 1$ then the Dirichlet probability density function is of the form $\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1}$, which is a β distribution (with $\alpha = \alpha_1$ and $\beta = \alpha_2$).

Note. The Dirac probability density function should not be confused with the Dirac *delta* distribution. See my online notes for Real Analysis 1 (MATH 5210) on Supplement. The Dirac Delta Function, A Cautionary Tale for more details.

Revised: 7/14/2021