STAT111: Mathematical Statisics **Spring 2025**

Week 2 Continuous Distributions: Gamma and Beta RV's

Instructor: Prof. Phil Everson Presenter: Zhengfei Li (Alex)

I would like to acknowledge that this handout is created with reference to Introduction to Probability by Blitzstein, J.K., and Hwang, J. (2014). I would also like to acknowledge the instructions from Prof. Everson, and some high level discussion with Cici Wen.

1 Preliminaries

1.1 Gamma distribution

Say we have random variable $X \sim \text{Gamma}(a, \lambda)$ with parameters $a > 0$ and $\lambda > 0$, then its PDF is:

$$f(x) = \frac{1}{\Gamma(a)}(\lambda x)^a e^{-\lambda x}\frac{1}{x}, \quad y > 0$$

Remark: While an Exponential r.v. represents the waiting time for the first success a Gamma r.v. represents the total waiting time for multiple successes.

Theorem. *If $X_i \overset{i.i.d.}{\sim}$ Gamma(a_i, λ) for $i = 1, \cdots, N$, then*

$$\sum_{i=1}^{N} X_i \sim Gamma(\sum_{i=1}^{N} a_i, \lambda)$$

To show this, it is sufficient to show that $X_1 + X_2 \sim \text{Gamma}(a_1 + a_2, \lambda)$. Then, one may show by induction that the theorem generalizes to the sum of N Gamma random variables.

Proof.

$$\begin{aligned}
f_{X_1+X_2}(s) &= \int_0^s f_{X_1}(x) f_{X_2}(s-x)dx \\
&= \int_0^s \frac{1}{\Gamma(a_1)}(\lambda x)^{a_1} e^{-\lambda x}\frac{1}{x} \cdot \frac{1}{\Gamma(a_2)}(\lambda(s-x))^{a_2} e^{-\lambda(s-x)}\frac{1}{s-x}dx \\
&= \int_0^s \frac{1}{\Gamma(a_1)\cdot\Gamma(a_2)}(\lambda)^{a_1+a_2} e^{-\lambda s} x^{a_1-1}(s-x)^{a_2-1}dx \\
&= e^{-\lambda s}(\lambda)^{a+b}\frac{1}{\Gamma(a_1+a_2)}\int_0^1 \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\cdot\Gamma(a_2)}(st)^{a_1-1}(s-st)^{a_2-1}sdt \\
&= e^{-\lambda s}(\lambda)^{a+b}\frac{1}{\Gamma(a_1+a_2)}\cdot s^{a_1+a_2-1}\int_0^1 \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\cdot\Gamma(a_2)}(t)^{a_1-1}(1-t)^{a_2-1}dt \\
&= \frac{e^{-\lambda s}(\lambda s)^{a+b}}{\Gamma(a_1+a_2)}\cdot\frac{1}{s}
\end{aligned}$$

Thus, we recognize that $X_1 + X_2 \sim \text{Gamma}(a_1 + a_2, \lambda)$. $\square$

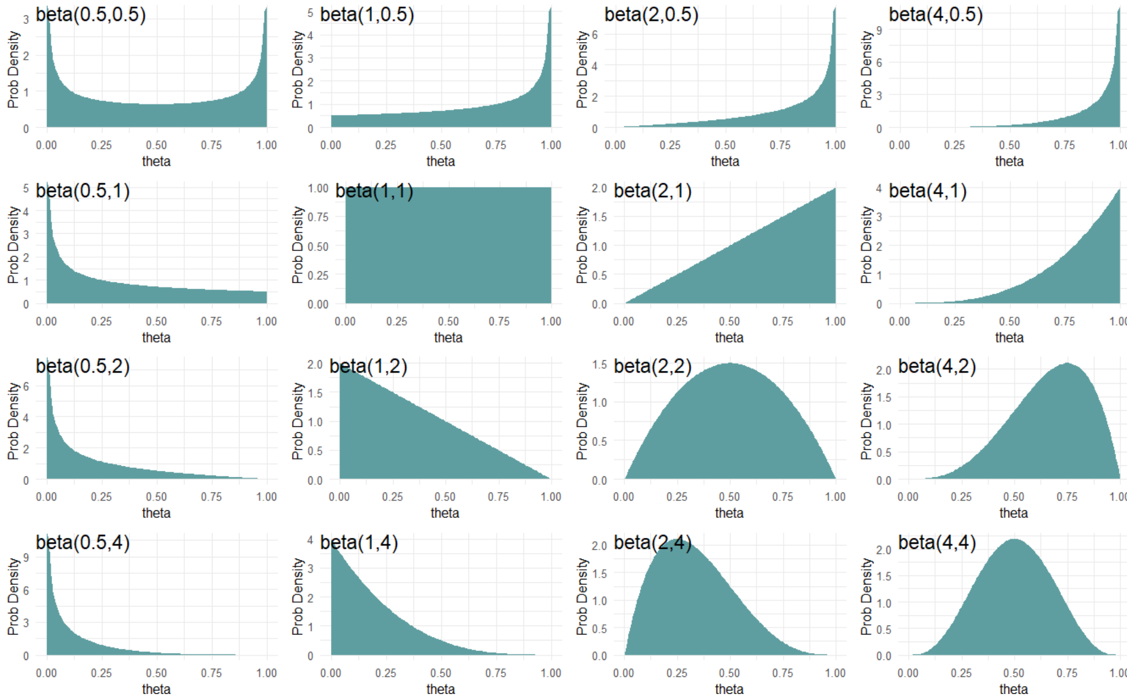

Figure 1: Beta Distribution with different parameters (source: www.causact.com)

1.2 Beta distribution

Say we have a random variable $X \sim Beta(a, b)$ with parameters $a > 0$ and $b > 0$, then its PDF is:

$$f(x) = \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1$$

Remark: Notice that Beta Distribution PDF has a zero or a singularity at $x = 0$ and $1 - x = 0 \Leftrightarrow x = 1$, we can then discuss the shape of Beta Distribution by separating cases according to the sign of $a - 1$ and $b - 1$.

Let's focus on the component x^{a-1} and we can make the following observations

- If $0 < a < 1$ and $x \to 0$, $x^{a-1} \to \infty$ and $\lim\limits_{x \to 0} f(x) \to \infty$.
- If $a = 1$ and $x = 0$, $x^{a-1} = 1$ and $f(0)$ would be a non-zero constant.
- If $a > 1$ and $x = 0$, $x^{a-1} = 0$ and $f(0)$ would be a non-zero constant.
- For x fixed in range $0 < x < 1$, x^{a-1} is larger for smaller a.

We can make similar observations to $(1 - x)^{b-1}$ component. Accordingly, we may visualize Beta Distribution with different values of parameters in Figure 1.

1.3 Other helpful facts

- Gamma function
  - $\Gamma(a + 1) = a\Gamma(a)$
  - $\Gamma(a) = (a - 1)!$ for $a \in \mathbb{Z}^+$

2 Problem 4

Let $V_1 \sim \text{Gamma}(a, \lambda)$ be independent of $V_2 \sim \text{Gamma}(b, \lambda)$. These variables may be used to define the Beta, F^* and F distributions.

2.1 Part (a)

Define $S = V_1 + V_2$ and $X = \frac{V_1}{V_1+V_2}$. Find the joint pdf for S and X and show that these are independent $\text{Gamma}(a+b, \lambda)$ and $\text{Beta}(a, b)$ random variables. Explain what this implies about the waiting time for some number of Poisson events, and the proportion of that time spent waiting for the first event (e.g.). Give the analogous explanation in terms of squared, centered Normal random variables.

2.1.1 Solution

As a general strategy, we are going to approach this problem through change of variable.

Proof. Given the distribution of Gamma, we know that independent random variables V_1 and V_2 has joint distribution:

$$\begin{aligned}
f_{V_1,V_2}(v_1, v_2) &= \frac{1}{\Gamma(a)}(\lambda v_1)^a e^{-\lambda v_1}\frac{1}{v_1} \cdot \frac{1}{\Gamma(b)}(\lambda v_2)^b e^{-\lambda v_2}\frac{1}{v_2} \\
&= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} v_1^{a-1} v_2^{b-1} e^{-\lambda(v_1+v_2)}
\end{aligned}$$

As we know that $S = V_1 + V_2$ and $X = \frac{V_1}{V_1+V_2}$, we know that through algebraic manipulation:

$$\begin{cases} V_1 = SX \\ V_2 = S(1-X) \end{cases}$$

Recall change of variables from several variable calculus:

$$\begin{aligned}
f_{S,X}(s, x) &= f_{V_1,V_2}(sx, s(1-x)) \cdot \det \begin{bmatrix} \frac{\partial v_1}{\partial s} & \frac{\partial v_2}{\partial s} \\ \frac{\partial v_1}{\partial x} & \frac{\partial v_2}{\partial x} \end{bmatrix} \\
&= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)}(sx)^{a-1}[s(1-x)]^{b-1} e^{-\lambda(sx+s(1-x))} \cdot \det \begin{bmatrix} x & s \\ 1-x & -s \end{bmatrix} \\
&= \left(\frac{\lambda^{a+b}}{\Gamma(a+b)} s^{a+b-1} \cdot e^{-\lambda s}\right) \cdot \left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\right)
\end{aligned}$$

Notice that the equation above can be separated to two components, respectively only dependent on S and X

By integral with respect to x ($f_S(s) = \int_0^1 f_{S,X}(s, x)dx$), we recognize that $f_S(s) = \frac{\lambda^{a+b}}{\Gamma(a+b)} s^{a+b-1} \cdot e^{-\lambda s}$, so $S \sim \text{Gamma}(a+b, \lambda)$. For support of S, we realize that $V_1, V_2 > 0$, so $S = V_1 + V_2 > 0$.

By integral with respect to s ($f_X(x) = \int_0^\infty f_{S,X}(s, x)ds$), we recognize that $f_X(x)\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$, so $X \sim \text{Beta}(a, b)$. For support of X, we realize that $V_1, V_2 > 0$, so $0 < X = \frac{V_1}{V_1+V_2} < 1$. $\square$

Remark: we can interpret V_1 as the waiting time for the first a events, and V_2 as the waiting time for the following b events. Then, S would be the total waiting time for $a+b$ events, and X would be the ratio of first waiting time to the overall waiting time. By our previous proof, we know that the total waiting time is independent of the fraction of time spent on the first a events.

Remark: Recall that in linear regression of responses on $p-1$ predictors and n observations, assuming normal error distribution. Under null hypothesis that $\beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$, we know that $\frac{SSM}{\sigma^2} \sim \chi^2_{p-1} = \text{Gamma}(\frac{p-1}{2}, \frac{1}{2})$ is independent of $\frac{SSE}{\sigma^2} \sim \chi^2_{n-p} = \text{Gamma}(\frac{n-p}{2}, \frac{1}{2})$. Then, $R^2 = \frac{SSM/\sigma^2}{(SSM+SSE)/\sigma^2} \sim \text{Beta}(\frac{p-1}{2}, \frac{n-p}{2})$.

2.2 Part (b)

Use integration by recognition (using the fact that pdf 's must integrate to 1) show that $E(X) = \frac{a}{a+b}$.

2.2.1 Solution

Proof. From (a), we know that $X \sim \text{Beta}(a, b)$. By formula of expectation:

$$
\begin{aligned}
E(X) &= \int_0^1 x \cdot x^{a-1}(1-x)^{b-1}dx \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \int_0^1 x^a(1-x)^{b-1}dx && \text{recognize kernel of Beta}(a+1, b) \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \cdot \int_0^1 \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)} x^a(1-x)^{b-1}dx \\
&= \frac{\Gamma(a+b)}{\Gamma(a+b+1)} \frac{\Gamma(a+1)}{\Gamma(a)} && \text{by definition of gamma function} \\
&= \frac{a}{a+b}
\end{aligned}
$$

□

2.3 Part (c)

Show that Uniform$(0, 1)$ is a special case of the Beta distribution. As an example of a Uniform variable, consider the joint pdf of V_1 and X, the times of the first and second events from a Poisson process ($V_1 < X$). Show that the conditional distribution of $V_1|X = x_0$ is Uniform by writing V_1 and $X = V1 + V2$, for V_1 and V_2 iid Exponential variables. Note how $V_1|X = x_0$ is Uniform, but for a constant y, the distribution of $V_1|V_1 < x_0$ is not Uniform.

2.3.1 Solution

We know that $V_1 \sim \text{Expo}(\lambda) = \text{Gamma}(1, \lambda)$ is independet of $V_2 \sim \text{Expo}(\lambda) = \text{Gamma}(1, \lambda)$. By conclusion following (a), we know that $X = \frac{V_1}{V_1+V_2} \sim \text{Beta}(1, 1) = \text{Uniform}(0, 1)$.

Fix the sum of $S = V_1 + V_2 = s_0$. By the distribution above, we know that $(V_1 = X \cdot S)|S = s_0 \sim \text{Uniform}(0, s_0)$.

Remark: Though $X = x_0$ suggests $V_1 < x_0$, we want to realize that $X = x_0$ is not a necessary condition for $V_1 < x_0$. Intuitively, we may think that only with $V_1 < x_0$, we did not exert constraints on the distribution for $x_0 - V_1$. Mathematically, we can think that $f(v_1|V_1 < x_0) = \frac{\lambda e^{-\lambda v_1}}{\int_0^t \lambda e^{-\lambda u}du}$, which is a segmented exponential distribution, clearly not being uniform. (draw a distribution diagram)

2.4 Part (d)

For $U_1, \cdots, U_n \overset{i.i.d.}{\sim}$ Uniform$(0,1)$, define $U(k)$ to be the kth order statistic, with $0 < U(1) < \cdots < U(n) < 1$. Use a differential argument to show that $U(k) \sim$ Beta$(k, n-k+1)$, for $k = 1, \cdots, n$.

2.4.1 Solution

Proof. By the definition of order statistics, we know that there would be $k-1$ random variables evaluated to be lower than $U(k)$, $U(k)$ to be in the small neighborhood where it is evaluated, and there would be $n-k$ random variables evaluated above $U(k)$. Thus, let's consider the probability of such event happening.

$$\begin{aligned}
f_{U(k)}(u)du &\approx P(U(k) \in [u, u+du]) \\
&= P(\text{k-1 of } u_i's < u) \cdot P(\text{one } u_i \in [u, u+du]) \cdot P(\text{n-k of } u_i's > u+du) \\
&= \binom{n}{k-1} u^{k-1} \cdot \binom{n-k+1}{1} du \cdot \binom{n-k}{n-k} (1-(u+du))^{n-k} \\
&= \frac{n!}{(n-k+1)!(k-1)!} \cdot (n-k+1) \cdot u^{k-1}(1-(u+du))^{n-k} \cdot du \\
&= \frac{n!}{(n-k)!(k-1)!} \cdot u^{k-1}(1-(u+du))^{n-k} \cdot du \\
\lim_{du \to 0} \frac{f_{U(k)}(u)du}{du} &= \frac{\Gamma(n+1)}{\Gamma(n-k+1)\Gamma(k)} \cdot u^{k-1}(1-u)^{n-k}
\end{aligned}$$

Thus, by recognition, $U(k) \sim Beta(k, n-k+1)$ with support $0 < U(k) < 1$.

By part (b), we also know that $E(U(k)) = \frac{k}{k+n-k+1} = \frac{k}{n+1}$. □

Remark: For example, if we have $n = 5$, the expected value of the ordered statistics are:

$$\begin{aligned}
E(U(1)) &= \frac{1}{6} \\
E(U(2)) &= \frac{2}{6} = \frac{1}{3} \\
E(U(3)) &= \frac{3}{6} = \frac{1}{2} \\
E(U(4)) &= \frac{4}{6} = \frac{2}{3} \\
E(U(5)) &= \frac{5}{6}
\end{aligned}$$

Along with "0" and "1", these numbers are uniformly distributed along the number range $[0, 1]$, which makes intuitive sense.