Stat 111 Week 2: Continuous Distributions

1. **Uniform random variables** (Blitzstein 5.2, 5.3)

   a) Give the probability mass function (pmf) for a discrete Uniform variable X, defined on the integers $1, \ldots, n$. Sketch the probability histogram for $n = 6$ (e.g., for rolling a six-sided die). Note how, with bars of width 1.0, the probability of any subset of the sample space (e.g., odd numbers) is equal to the area of the bars corresponding to those outcomes. This means the pmf is also a *density* function.

   b) Define $U = X/n$, with $U \in (0, 1]$ for all n. To have the area of the bars correspond to probability for U, divide the pmf by the width of each bar to get the density function $f_u(u)$. Describe the limit of this probability density function (pdf) and of this discrete random variable as $n \to \infty$. Define the continuous Uniform random variable $U \sim \text{Unif}(0, 1)$ and the more general $\text{Unif}(a, b)$ distribution.

   c) Let F be the cdf for a continuous random variable, with F a monotone increasing function (briefly explain the other possibilities). Let $G(u)$ be the inverse function, so that $G(u) = x$ if $F(x) = u$. For a random variable X that has cdf F, show that $U = F(X)$ follows a $\text{Unif}(0, 1)$ distribution. Give intuition for why this makes sense.

   d) Reversing the calculation from part d, suppose $U \sim \text{Unif}(0, 1)$ and define the transformed random variable $X = G(U)$. Show that X has cdf F. This is called the "inverse-cdf method" for random number generation. As an example, for $U \sim \text{Unif}(0, 1)$ and a constant $\lambda > 0$, show that $X = -\log(U)/\lambda$ follows an Exponential(λ) distribution with cdf $F(x) = 1 - e^{-\lambda x}$ for $x > 0$. Compare a large sample generated using this method to a sample generated using rgamma. Compare the moments and plot the sorted values of each sample against each other to see how close they come to the line $y = x$.

2. **The Differential Argument and the Gamma Distribution** (Blitzstein 5.5, 5.6, 13)
   The differential approach to finding a pdf $f_x(x)$ is summarized as follows:

   i) Write out an expression for $P(X \in [x, x + dx))$ in terms of x and a small positive increment dx, and equate this with $f_x(x)dx$, the pdf evaluated at x, multiplied by dx.

   ii) Remove terms that go to 1 in the limit as $dx \to 0$ (e.g., e^{cdx} for any constant c).

   iii) Divide both sides by dx, leaving $f_x(x)$ on one side. Take the limit as $dx \to 0$ to find the density function.

   a) For a continuous random variable X with pdf f_x, explain why, for a small deviation dx,

   $$P(X \in [x, x + dx)) \approx f_x(x)dx$$

   Show how this becomes an exact equivalence when you divide both sides by dx and let dx approach 0 (or if the density $f_x(x)$ is a constant function). As an example, derive the Exponential density for X by finding the probability the first Poisson event occurs in the interval $[x, x + dx)$. Then divide by dx and take the limit $dx \to 0$. Show this agrees with the result you get by finding the Exponential cdf and differentiating.

b) Use a differential argument to derive the Gamma(k, λ) density. Let Y be the time of the kth event for a Poisson process with rate λ. Find the probability of $k-1$ events occurring before time y and a kth event occurring in the interval $[y, y+dy)$. Divide by dy and take the limit as $dy \to 0$. Give intuition for why, if $X_1 \sim \text{Gamma}(k_1, \lambda)$ is independent of $X_2 \sim \text{Gamma}(k_2, \lambda)$, then $X_1 + X_2 \sim \text{Gamma}(k_1 + k_2, \lambda)$, and for why the sum is not Gamma if the λ's differ.

c) For a constant $c > 0$, show that if $X \sim \text{Gamma}(\alpha, \lambda)$, then $Y = cX \sim \text{Gamma}(\alpha, \lambda/c)$. Give intuition based on units of measure. For example, consider X to be a waiting time in hours and $Y = 60X$ to be the time in minutes.

d) Define the Gamma function and show how it normalizes the Gamma(α, λ) density for any value $\alpha > 0$. Give the recursive property of the Gamma function and the connection to the factorial function.

3. **The Normal and Chi-square Distributions** (Blitzstein 5.4, 8, Rice 6.2)

a) Show that the square of a standard Normal variable follows a Gamma$(1/2, 1/2)$ distribution (and that $\Gamma(1/2) = \sqrt{\pi}$). This distribution is also called $\chi^2_{(1)}$, or Chi-square with 1 degree of freedom. The sum of k independent squared standard Normal variables is $\chi^2_{(k)}$. Use facts about the Gamma distribution to show that if $X \sim \chi^2_{(\nu)}$, then $X \sim$Gamma$(\frac{\nu}{2}, \frac{1}{2})$. Also show that if $Y \sim$Gamma(α, λ), then $X = 2\lambda Y \sim \chi^2_{(2\alpha)}$. For example, show that $X \sim N(0, \sigma^2)$ implies $X^2 \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2\sigma^2})$ and $X^2/\sigma^2 \sim \chi^2_{(1)}$.

b) For $Z_1, Z_2 \overset{\text{i.i.d.}}{\sim} N(0, 1)$, derive the pdf for $R = \sqrt{Z_1^2 + Z_2^2}$ and $\theta = \text{atan}(Z_2/Z_1)$, the polar coordinates for the point (Z_1, Z_2). R is the distance from the point to the origin and θ is the angle made between the vector through (Z_1, Z_2) and the z_1-axis. Explain the implications of this joint pdf being constant in θ.

c) For two independent standard Normal coordinates Z_1 and Z_2, note how the squared distance of the point (Z_1, Z_2) from the origin is an Exponential$(1/2)$ random variable (because $Z_1^2 + Z_2^2 \sim \chi^2_{(2)}$, by definition). Also show that the area A of a circle centered at the origin and passing through the point (Z_1, Z_2) is an Exponential random variable. This is the amount of area you can you sweep out before you encounter an "event", and is analogous to "waiting time" for a Poisson process in the plane.

d) Show how two independent Uniform variables may be transformed to generate two independent $N(0, 1)$ random variables (the Normal inverse cdf is not available in closed form).

4. **Gamma and Beta Random Variables** (Blitzstein 8.3-8.5, Rice 2.2, 3.7)
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.

a) Define $S = V_1 + V_2$ and $X = V_1/(V_1 + V_2)$. Find the joint pdf for S and X and show that these are independent Gamma$(a + b, \lambda)$ and 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.

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

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 X and Y, the times of the first and second events from a Poisson process ($X < Y$). Show that the conditional distribution of $X|Y = y$ is Uniform by writing $X = V_1$ and $Y = V_1 + V_2$, for V_1 and V_2 iid Exponential variables. Note how $X|Y = y$ is Uniform, but for a constant y, the distribution of $X|X < y$ is *not* Uniform.

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

5. ***F* and *t* Random Variables** (Rice 6.2)

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

a) Let $Y = c\frac{V_1}{V_2}$, for a constant $c > 0$. Then Y follows what I call the $F^*(a, b, c)$ distribution ("F-star", with Gamma convolution parameters a and b, and multiplicative constant c). Show using representation (i.e., using the definition) that the ratio of any two independent Gamma (or Chi-square) random variables follows an F^* distribution.

b) For $X = \frac{V_1}{V_1+V_2} \sim$ Beta(a, b), show using representation that $R = \frac{X}{1-X} \sim F^*(a, b, 1)$. Use this fact to derive the pdf for R and for $Y = cR$.

c) The usual $F_{(m_1,m_2)}$ distribution is defined in terms of two independent Chi-square variables. If $a = m_1/2$, $b = m_2/2$ and $\lambda = 1/2$, then $V_1 \sim \chi^2_{(m_1)}$ independent of $V_2 \sim \chi^2_{(m_2)}$, and $Y = (V_1/m_1)/(V_2/m_2) \sim F_{(m_1,m_2)}$. Show that Y also follows an F^* distribution.

d) If $Z \sim N(0,1)$ is independent of $V \sim \chi^2_{(m)}$, then $T = \frac{Z}{\sqrt{V/m}} \sim t_{(m)}$, the t distribution with m degrees of freedom. Show that T^2 is an F variable. Use this fact (and symmetry) to derive the $t_{(m)}$ pdf.