Wednesday, October 25 **Fall 2023**

Section 6: Universality of the Uniform, Normal, Expo, and Moments

Srihari Ganesh

Based on section note formatting template by Rachel Li and Ginnie Ma '23

Forms

- Attendance form: http://bit.ly/110attend
- Feedback form: https://bit.ly/SrihariFeedback

Attendance form

Section feedback form

1 Summary

No MGF problems on the pset this week.

1.1 Universality of the Uniform

Recall that the standard uniform, $U \sim \text{Unif}(0, 1)$, has support $(0, 1)$ with PDF 1 in the support.

Theorem 1 (Universality of the Uniform, UoU)**.** *If F is a valid CDF that is continuous and strictly increasing over the support, then*

1. *Let $U \sim Unif(0, 1)$. Then $F^{-1}(U)$ is a random variable with CDF F.*
2. *Let X have CDF F. Then $F(X) \sim Unif(0, 1)$.*

The first result applies to discrete random variables as well. The second result only works for continuous random variables.

Proof. For continuous random variables with F as described in the theorem,

1. For $x \in \mathbb{R}$,

$$P(F^{-1}(U) < x) = P(F(F^{-1}(U)) < F(x)) = P(U < F(x)) = F(x).$$

So $F^{-1}(U)$ has CDF F. We used the CDF of U in the last step, since $F(x) \in [0,1]$.

2. For $u \in [0,1]$,

$$P(F(X) < u) = P(F^{-1}(F(X)) < F^{-1}(u)) = P(X < F^{-1}(u)) = F(F^{-1}(u)) = u,$$

so $F(X) \sim \text{Unif}(0,1)$ since it has the CDF of a standard uniform.

$\square$

1.2 Normal distribution

Definition 2 (Standard Normal)**.** $Z \sim \mathcal{N}(0,1)$ is a **standard Normal** random variable with support $\mathbb{R}$. We notate the CDF as Φ and PDF as ϕ.

Result 3 (Symmetry)**.** The standard Normal is symmetric about 0. In math, for $x \in \mathbb{R}$, $\phi(x) = \phi(-x)$.

- This also implies that $\Phi(x) = 1 - \Phi(-x)$.
  - So $\Phi(0) = 0.5$.
- For $Z \sim \mathcal{N}(0,1)$, $-Z \sim \mathcal{N}(0,1)$ as well.

Result 4 (Empirical rule/68-95-99.7 rule)**.**

$$\begin{aligned} P(-1 < Z < 1) &\approx 0.68, \\ P(-2 < Z < 2) &\approx 0.95, \\ P(-3 < Z < 3) &\approx 0.997. \end{aligned}$$

In this class, you can give exact answers in terms of Φ and ϕ. On psets, you should also use a calculator/programming language/the empirical rule to get numerical approximations of Φ.

Definition 5 (General Normal)**.** $X \sim \mathcal{N}(\mu, \sigma^2)$ (with $\mu \in \mathbb{R}, \sigma > 0$) is a **Normal** random variable with mean μ and variance σ^2, and also has support $\mathbb{R}$.

Result 6 (Location-scale)**.** For $Z \sim \mathcal{N}(0,1)$, $\mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$.
More generally, for $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $\mu_2 + \sigma_2 X \sim \mathcal{N}(\mu_2 + \mu_1\sigma_2, \sigma_1^2\sigma_2^2)$.

Result 7 (Standardization)**.** For $X \sim \mathcal{N}(\mu, \sigma^2)$, $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.
We often use this to get results in terms of Φ:

$$P(X < x) = P(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}) = \Phi(\frac{x-\mu}{\sigma}).$$

Corollary 8 (Empirical rule)**.** *For* $X \sim \mathcal{N}(\mu, \sigma^2)$,

$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &\approx 0.68 \\ P(\mu - 2\sigma < X < \mu + 2\sigma) &\approx 0.95 \\ P(\mu - 3\sigma < X < \mu + 3\sigma) &\approx 0.997 \end{aligned}$$

Result 9 (Sum of independent Normals)**.** Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ with X, Y independent. Then

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2),$$
$$X - Y \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2).$$

☣ **10** (Variance when subtracting)**.** See that we always add the variance above! This is also a general rule: for any independent random variables X and Y,

$$Var(X + Y) = Var(X - Y) = Var(X) + Var(Y).$$

1.3 Exponential distribution

Definition 11 (Exponential distribution)**.** $X \sim \text{Expo}(\lambda)$ is an **Exponential** random variable with mean $\frac{1}{\lambda}$ and variance $\frac{1}{\lambda^2}$. λ is called the **rate parameter**.

Result 12 (Memorylessness)**.** For $X \sim \text{Expo}(\lambda)$ and any $s, t > 0$, the **memoryless** property of the Exponential distribution states the following (equivalent) results:

$$P(X > s + t | X > s) = P(X > t)$$
$$(X - s | X > s) \sim \text{Expo}(\lambda).$$

See specifically that $X - s | X > s$ is independent of the value of s.

The Exponential distribution is the only continuous distribution with this property. Additionally, the Geometric distribution is the only discrete distribution with support $\{0, \ldots, \}$ that is memoryless.

☣ **13.** For most results we talk about, you can't put a random variable in the place of a constant - you might recall from last week's problem set that we couldn't let the sum of N independent Pois(λ) r.v.s, with N random, be distributed Pois($N\lambda$). However, with memorylessness, you can put random variables in the place of the s above - so for some random variable Y, $(X - Y | X > Y) \sim \text{Expo}(\lambda)$ still.

Example 14 (Memorylessness)**.** Suppose you're waiting for a bus that will arrive in $X \sim \text{Expo}(\lambda)$ minutes. If you wait for the bus for 10 minutes and it has not arrived, then the remaining time that you have to wait is still distributed Expo(λ): $X - 10 | X > 10 \sim \text{Expo}(\lambda)$. So no matter how long you wait, the remaining time for you to wait has the same distribution.

Result 15 (Minimum of Expos)**.** The minimum of n i.i.d. Expo(λ) random variables is distributed Expo($n\lambda$). In notation, for $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Expo}(\lambda)$, $\min(X_1, \ldots, X_n) \sim \text{Expo}(n\lambda)$.

☣ **16** (Maximum of Expos)**.** The maximum of n i.i.d. Exponential distributions is *not* does not have an Exponential distribution.

Remark 17 (Finding the distribution of minimums/maximums)**.** The proofs for the results above can be found in the book, but they provide a general template for finding the distributions of minimums and maximums.

Let $X_1, \ldots, X_n$ be any random variables. Then the events $\{\min(X_1, \ldots, X_n) > x\}$ and $(X_1 > x) \cap (X_2 > x) \cap \cdots \cap (X_n > x)$ are equivalent. To convince yourself of this, think about what this means in words: the minimum of a set of numbers is greater than x if and only if each one of the numbers is great than x.

To find the CDF of $\min(X_1, \dots, X_n)$, a common workflow is

$$P(\min(X_1, \dots, X_n) \leq x) = 1 - P(\min(X_1, \dots, X_n) > x) = 1 - P(X_1 > x, X_2 > x, \dots, X_n > x).$$

If $X_1, \dots, X_n$ are independent, then we can get that

$$P(X_1 > x, X_2 > x, \dots, X_n > x) = P(X_1 > x)P(X_2 > x) \cdots P(X_n > x)$$

If $X_1, \dots, X_n$ are also identically distributed, we conclude with

$$P(X_1 > x)P(X_2 > x) \cdots P(X_n > x) = (P(X_1 > x))^n.$$

For maximums, we follow a similar workflow, except instead using the fact that

$$\{\max(X_1, \dots, X_n) < x\} = \bigcap_{i=1}^{n} (X_i < x).$$

1.4 Moments/Moment Generating Functions

Definition 18 (Moments)**.** For a random variable X, the $\mathbf{n^{th}}$ **moment** is $E(X^n)$.

Definition 19 (Moment Generating Function)**.** For a random variable X, the **moment generating function (MGF)** is $M_X(t) = E(e^{tX})$ for $t \in \mathbb{R}$. If the MGF exists, then

$$M_X(0) = 1,$$
$$\frac{d^n}{dt^n} M_X(t)|_{t=0} = M_X^{(n)}(t) = E(X^n).$$

You should sanity-check that $M_X(0) = 1$ whenever you calculate an MGF.

2 Practice Problems

1. Xavier and Youssef are running a 10K race. Xavier's time (in minutes) is $X \sim \mathcal{N}(50, 3^2)$, while Youssef's time is $Y \sim \mathcal{N}(52, 4^2)$. Their times are independent.

   (a) What is the probability that Youssef runs the 5K in under an hour? Answer in terms of Φ.

   (b) Use the empirical rule to give a simple numerical approximation for your answer to (a).

(c) What is the probability that Xavier beats Youssef by at least a minute? Give your answer in terms of Φ.

(d) What is the probability that Xavier beats Youssef by at least two minutes? Give an exact answer.

2. This problem is meant to develop a strong base to do Problem 5 on this week's problem set. Let $T_1, T_2 \overset{i.i.d.}{\sim} \text{Expo}(\lambda)$ be the times it takes for two radioactive particles to decay. Define $M = \max(T_1, T_2)$.

   (a) Find the CDF of M. *Hint: use the strategy from remark 17.*

   (b) Express M as the sum of two Expo random variables, and find the rate parameters for each of those random variables. *Hint: use both memorylessness (Result 12) and the distribution of the minimum of Expos (Result 15).*