Probability Theory

Textbook: *Introduction to Probability* by Blitzstein and Hwang

Previous Lecture

- Poisson
- Continuous Rv: PF/CDF
- Cont. Valid PFs
- Cont. Expectation/LOTUS

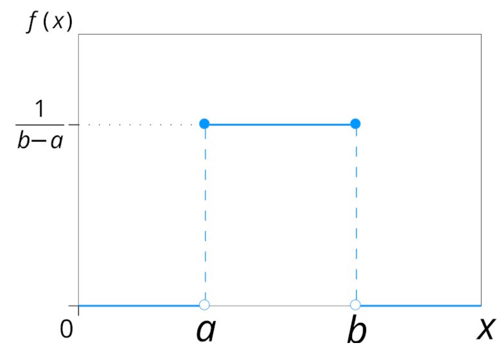
§5.2 - Uniform Distr

A Uniform rv on (a, b) is a simple random number between a and b.

Def (**Uniform Distr**): A cont U has Uniform distr on interval (a, b) if its PF is:

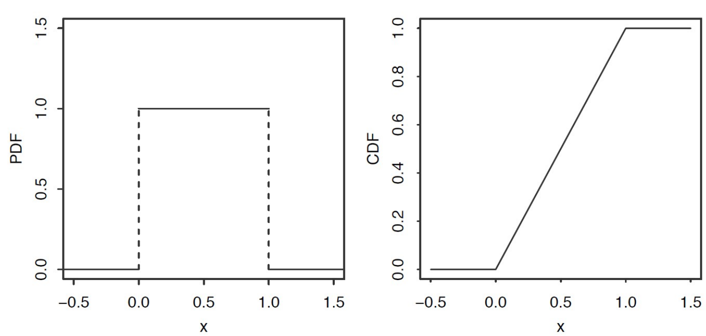
$$f(x) = \begin{cases} \frac{1}{b-a} \text{ if } a < x < b, \\ 0 \quad \text{otherwise.} \end{cases}$$

We denote this by $U \sim Unif(a, b)$.



The CDF is the accumulated area under the PF: $F(x) = \begin{cases} 0 \text{ if } x \leq a \\ \frac{x-a}{b-a} \quad \text{if } a < x < b, \\ 1 \text{ if } x \geq b. \end{cases}$

Def (**Standard Uniform**): $Unif(0, 1)$.



$Unif(0, 1)$: PF and CDF

Ex (**Uniform Second Moment**): Let $X \sim Unif(a, b)$. Find $E(X^2)$.

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx$$

$$= \frac{1}{b-a} \left[\frac{1}{3} x^3 \right]_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{1}{3}(a^2 + b^2 + ab).$$ □

Proposition (**Prob is Proportionate to Length**). Let $U \sim Unif(a, b)$, and let (c, d) be a subinterval of (a, b), of length ℓ (so $\ell = d - c$). Then prob of U being in (c, d) is proportional to ℓ. A subinterval that's twice as long has twice the prob of containing U.

Proof. Since the PF of U is $\frac{1}{b-a}$ on (a,b), the area under the PF from c to d is $\frac{\ell}{b-a}$, which is a constant times ℓ. ■

Proposition (**Cond Distr of Unif is Unif**). Let $U \sim Unif(a,b)$, and let (c,d) be a subinterval of (a,b). Then the conditional distr of U given $U \in (c,d)$ is $Unif(c,d)$.

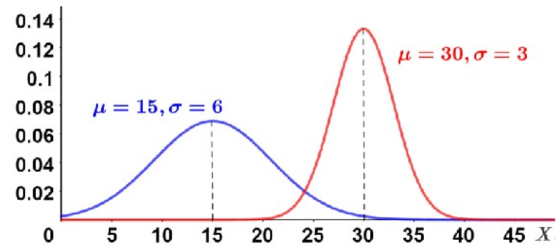
Proof. For u in (c,d), the conditional CDF at u is

$$P(U \le u | U \in (c,d)) = \frac{P(U \le u,\ c<U<d)}{P(U \in (c,d))} = \frac{P(U \in (c,u])}{P(U \in (c,d))} = \frac{\frac{u-c}{b-a}}{\frac{d-c}{b-a}} = \frac{u-c}{d-c}.$$

The conditional CDF is 0 for $u \le c$ and 1 for $u \ge d$. So the conditional distr of U is as claimed. ■

Location-Scale Transformation

Def (**Location-Scale Transformation**): Let $Y = \sigma X + \mu$, where σ and μ are constants w\$\sigma > 0$. Y is a *location-scale transformation* of X.
Here μ controls how location is changed, σ controls how scale is changed.



Transforming a Uniform rv this ways produces another Univorm rv.

Also, note: $E(Y) = E(\sigma X + \mu) = \sigma E(X) + \mu$. And $Var(Y) = \sigma^2 Var(X)$.

Ex (**Uniformly Distr Boats**): Suppose boats flow past your house on the river, and arrive in a uniformly distributed way between the 3rd and the 12th of the month. What's the prob that a particular boat coming down the river arrives between the 7th and the 11th?

Solution: They are uniformly distributed over the 10 days (including the 3rd & 12th): $U(3,13)$.

There are 5 days between the 7th and the 11th.

So $P(7 < x < 12) = P(x < 12) - P(x < 7) = F(12) - F(7) = \frac{12-3}{10} - \frac{7-3}{10} = \frac{5}{10} = 0.5$. □

[Next Time I Teach: Add the expectation in variance of a uniform distribution]

Harvard Video: https://www.youtube.com/watch?v=9vp1Ll2NpRw&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=15

§5.3 - Universality of the Uniform

Can we do transformations which take us from the Uniform distr to every other distr?

And can we transform non-uniform distrs into the Uniform distr? (yes)

Thm (**Universality of the Uniform**): Let F be a cont CDF which is *strictly* increasing on the support of the distr. This ensures the inverse F^{-1} exists from $(0,1) \to \mathbb{R}$. We then have:

1. Let $U \sim Unif(0,1)$ and $X := F^{-1}(U)$. Then X has CDF F.

2. Let X be a rv w/CDF F. Then $F(X) \sim Unif(0,1)$. (substituting X into its own CDF gives $Unif(0,1)$)

Proof:

1. Let $U \sim Unif(0,1)$ and $X = F^{-1}(U)$. For all real x, we need to show that $P(X \le x) = F(x)$.

   Observe that $P(X \le x) = P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x)$, so the CDF of X is F, as claimed.

   For the last equality, we used the fact that for $Unif(0,1)$, we have $P(U \le u) = u$ for $u \in (0,1)$.

2. Let X have CDF F. We want to show that $F(X) \sim Unif(0,1)$.

   Since F (being a CDF) takes values in $(0,1)$, then $P(F(X) \le y) = 0$ for $y \le 0$ and equals 1 for $y \ge 1$.

   And for $y \in (0,1)$, we have $P(F(X) \le y) = P(X \le F^{-1}(y))$

   $= F(F^{-1}(y))$ (since X has CDF F)

   $= y.$

   Thus $F(X)$ has the $Unif(0,1)$ CDF. ■

Visualization of Intuition: Youtube.com/watch?v=TzKANDzAXnQ

Ex (**Percentiles**): An exam is graded 0 to 100. Let X be the score of a random student. Let's approximate the discrete distr of scores w/a cont distr. So cont X has a CDF F that's strictly increasing on $(0,100)$. Suppose the median score is 60 (half of students > 60): $F(60) = \frac{1}{2}$; or equivalently, $F^{-1}(\frac{1}{2}) = 60$. If Fred gets 72, then his *percentile* is the fraction of students scoring below 72. So $F(72)$ is in $(\frac{1}{2}, 1)$.

In general, a student w/score x has percentile $F(x)$. If we start w/a percentile, 0.95, then $F^{-1}(0.95)$ is a score that has that percentile, or **quantile**. So, F^{-1} is called the **quantile function**.

The choice of plugging X into its own CDF has a natural interpretation: $F(X)$ is a percentile attained by a random student. The distr of scores on an exam are usually non-uniform. On the other hand, the distr of percentiles of students is uniform: the universality property says $F(X) \sim Unif(0,1)$.

Example: 50% of students have a percentile of at least 0.5. Universality of the Uniform says that 10% of students have a percentile between 0 and 0.1, 10% have a percentile between 0.1 and 0.2, 10% have a percentile between 0.2 and 0.3, and so on - a fact that is clear from the definition of percentile. □

Ex: Let X have PF: $f(x) = \begin{cases} \frac{x^3}{156} & 1 \le x \le 5, \\ 0 & \text{otherwise.} \end{cases}$

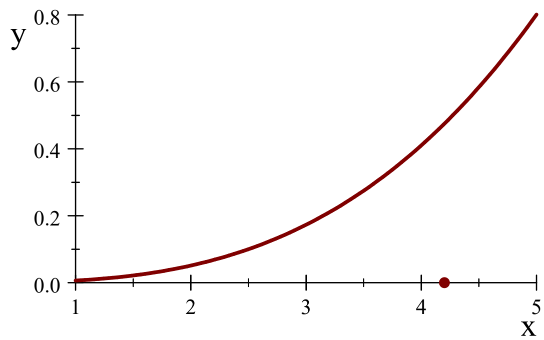
a. Find the median of X.

Solution: To use the Universality of the Uniform, we need the CDF.

$$F(x) = \int_1^x f(t)dt = \frac{1}{156}\int_1^x x^3 dx = \frac{1}{156}\left[\frac{1}{4}x^4\right]_1^x = \frac{x^4-1}{624}.$$

To find $F^{-1}(u)$, we set: $u = \frac{x^4-1}{624} \Rightarrow 624u+1 = x^4 \Rightarrow x = \sqrt[4]{624u+1}$. (positive since $1 \le x \le 5$)

So $F^{-1}(u) = \sqrt[4]{624u+1}$. And the median is: $F^{-1}(\frac{1}{2}) = \sqrt[4]{\frac{624}{2}+1} \approx 4.2$.



$f(x) = \frac{x^3}{156}$. Center of gravity is around 4.2

b. Find the 25th percentile of X.

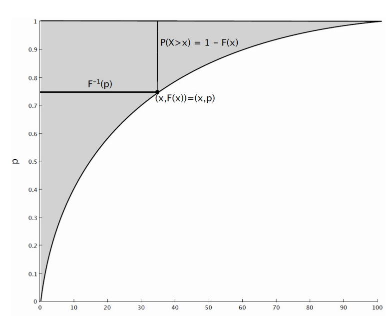
$F^{-1}(\frac{1}{4}) = \sqrt[4]{\frac{624}{4}+1} \approx 3.54$. □

Def (**Survival Function**). The survival function of X w/CDF F is $G(x) = 1 - F(x) = P(X > x)$. (notice "greater than")

Thm (**Expectation by Integrating the Survival**).
Let X be nonnegative. Its expectation can be found by integrating its survival: $E(X) = \int_0^\infty P(X > x)dx$.

[short proof in book]

The area above a certain CDF and below the line $p = 1$ is shaded.
This area can be interpreted in two ways: as the integral of the survival function,
or as the integral of the quantile function.



What did we learn?

- Uniform Distr
- Location-Scale Transformation
- Universality of the Uniform: Percentiles