Stat 135 <u>lec 25</u>

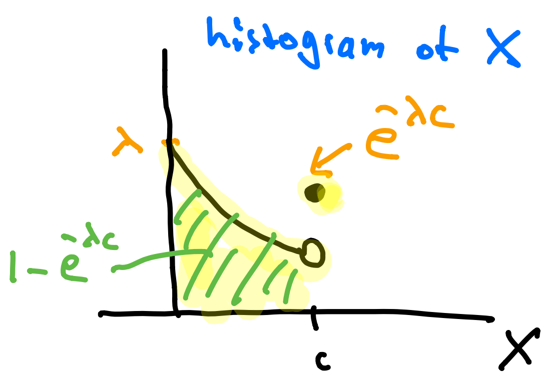
Quiz 3 sec 4.1, 4.2, 4.4, 4.5 (no 4.6 on Quiz 3 !)

sec 4.5 <u>cdf</u>

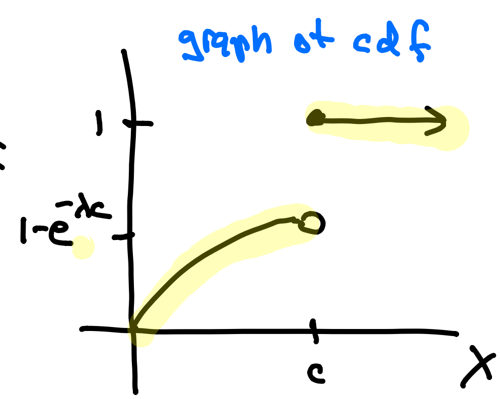
Last time :

$T \sim \text{expon}(\lambda)$, $c > 0$

$X = \min(T, c)$ called "T killed by c"



has cdf



mixed distribution

mixed cdf.

CDF has 3 main purposes

① Specify a distribution regardless of whether it is discrete, mixed, continuous.

ex X is RV w/ cdf $F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x/2} & \text{if } 0 \le x < 10 \\ 1 & \text{if } x \ge 10 \end{cases}$

specifies $X = \min(T, 10)$

for $T \sim \text{expon}(1/2)$ a mixed distribution.

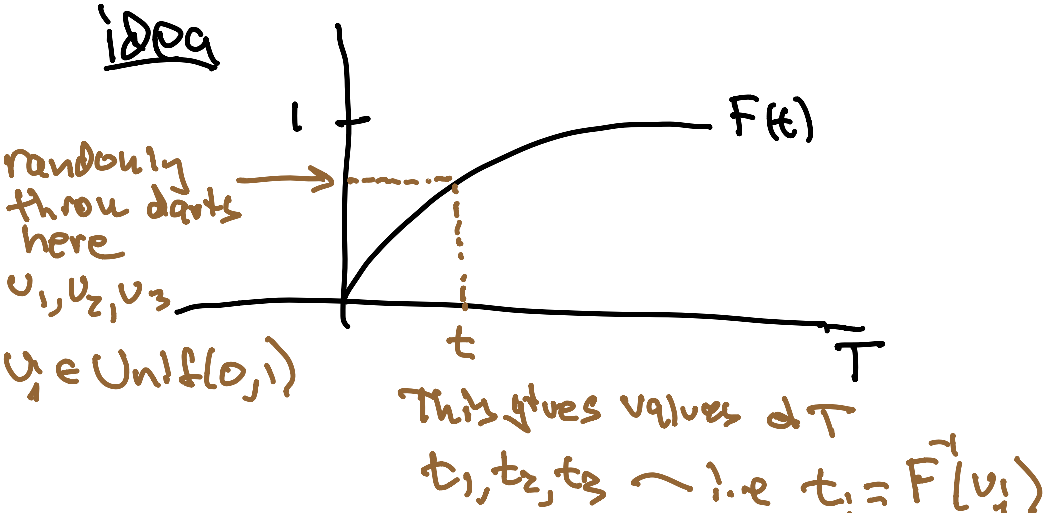
② Simulation of RV using the inverse cdf

Given a distribution on the line, use the cdf to generate a RV with that cdf

(namely if F is continuous, $X = F^{-1}(U)$ where $U = \text{Unif}(0,1)$)

This is how computers generate random numbers of any distribution for the purpose of simulation.

<u>idea</u>



explicitly, let $F(t) = 1 - e^{-\lambda t}$

$u = F(t) = 1 - e^{-\lambda t}$ for some t

solve for t

$e^{-\lambda t} = 1 - u$

$-\lambda t = \log_e (1-u)$

$$t = \frac{-\log_e(1-u)}{\lambda} = F^{-1}(u)$$

If $u \in \text{Unif}(0,1)$ how do we know that that $F^{-1}(u) \in T$? (i.e how do we know that $t_1, t_2, t_3, \ldots$ will have histogram of expon(λ)?)

See proof below:

See p 322 in book

<u>Claim</u> for any continuous cdf F,

$X = F^{-1}(U)$ is a RV with cdf F.

Unif(0,1)

Proof/ let $X = F^{-1}(U)$

$$
\begin{aligned}
F_X(x) &= P(X \le x) \\
&= P(F^{-1}(U) \le x) \\
&= P(F \cdot F^{-1}(U) \le F(x)) \\
&= P(U \le F(x)) \\
&= F(x)
\end{aligned}
$$

we will show $F_X = F$

since F is increasing

since $P(U \le u) = u$

$\square$

This is a little abstract so lets actually check by picking 100 random numbers $u_1, u_2, \ldots, u_{100}$ between 0 and 1 and seeing that the mean of $t_1, \ldots, t_{100}$ is $\frac{1}{\lambda}$ and SD of $t_1, \ldots, t_{100}$ is $\frac{1}{\lambda}$ (suggesting that $t_1, \ldots, t_{100}$ have dist. T)

In class demonstration in R:

ex Lets simulate lifetime of a bulb w/ rate of burnout $\lambda = 2$ bulbs/yr

i.e $T \sim \text{expon}(\lambda = 2)$

runif(100) in R

a vector of 100 random numbers between 0 and 1

recall $t = \dfrac{-\log_e(1-u)}{2}$

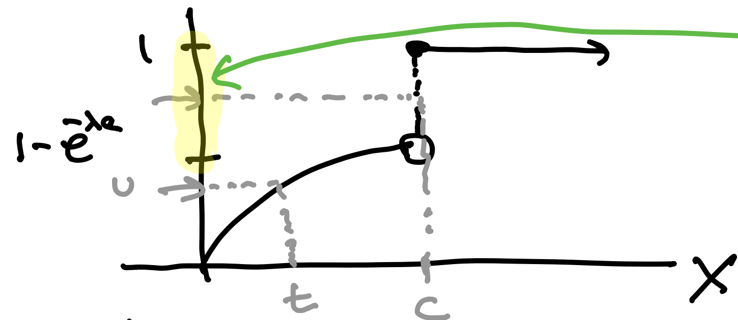
a vector of 100 numbers

lets check that mean $t \approx \frac{1}{2}$ ✓

sd $t \approx 1/2$. ✓

We can similarly do this for discrete or mixed distributions.

eg $X = \min(T, c)$, $T \sim \exp(\lambda)$



This yellow region gets assigned the single value c.

check:

$$P(F^{-1}(u) = c) = P(1-e^{-\lambda c} < u < 1)$$

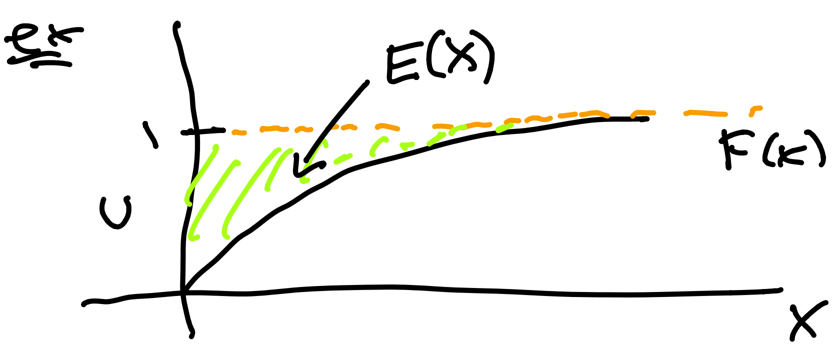
$$= 1 - (1-e^{-\lambda c}) = e^{-\lambda c}$$ as required ✓

③ <u>Expectation from cdf</u> (see 4.5.9)

It is sometimes easier to calculate $E(X)$ using the cdf (avoid doing integration by parts)

Let X be a pos random variable, with cdf F (cont, mixed, discrete)

eg

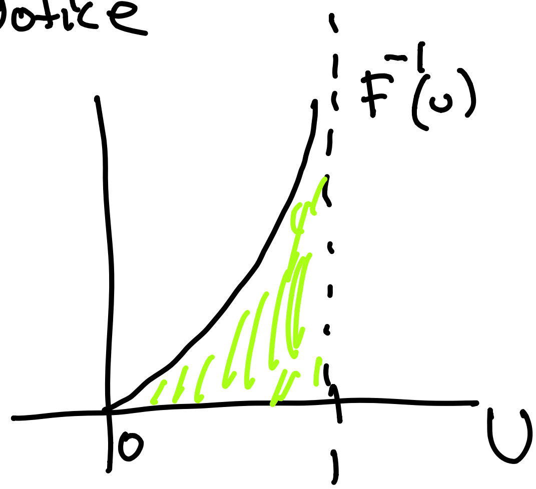


Claim: $E(X)$ is the shaded region above

$$E(X) = E\left(F^{-1}(U)\right) = \int_0^1 F^{-1}(u) f_U(u)\, du$$

$f_U(u) = 1$

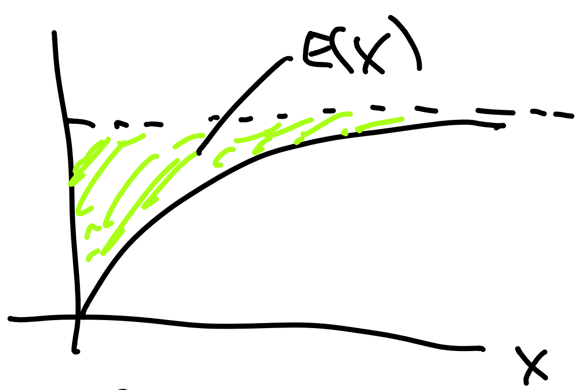
Notice



Here we reflect the above graph about the diagonal $y=x$

Indeed, the shaded region is $\int_0^1 F^{-1}(u)\, du$.

So, $E(X)$ is the shaded region below.



We can find the shaded region by integrating $1-F(x)$ with respect to x:

$$\boxed{E(X) = \int_0^\infty (1 - F(x))\,dx}$$

Notice this is equivalent to

$$E(X) = \int_0^\infty P(X > x)\,dx$$

which is the continuous version of the tail sum formula for expectation!

<u>ex</u>

$T \sim \text{expon}(\lambda)$

$F_T(t) = 1 - e^{-\lambda t}$

Calculate $E(T)$.

$$E(T) = \int_0^\infty (1 - F(t))\,dt = \int_0^\infty e^{-\lambda t}\,dt$$

$$= \frac{1}{\lambda} e^{-\lambda t}\Big|_0^\infty = \frac{1}{\lambda}$$

wow that was easy!